

What happened to the world's money?

Disappointment among financiers about the slow pace of the West's recovery from recession is understandable, but they should not be so surprised; who but they are helping to reinvest the world's capital internationally?

THE past few months have had the world's financiers gnashing their teeth, and for good reason. Their worry seems to be that the rules of the game they have been playing profitably for several decades seem to have changed. The recession that began in 1991 appears to be ending, but not according to the book. Both in Britain and the United States, for example, output appears to have reached a minimum a year ago, but since then its growth has been much slower than had been expected. The classical function of a recession, which is to impose such financial strains on economically marginal enterprises that they go bankrupt, appears to have worked all too efficiently; the bankruptcy courts have been thronged for a long time now. But the ending of the recession has not this time seemed like the liberation of productive enterprises, made lean and efficient during recession, from the financial restraints of high interest rates. What has gone wrong?

For one thing, although enterprises that managed to survive the recession may now be more efficient, they are often embedded in economies not in spectacularly good shape. The US federal deficit is still with us, but now the British government (which is running up debts at the rate of £50 billion a year), has one of proportionately comparable magnitude and can boast of a £20 billion annual trade deficit as well. Similarly, Germany's structural problems persist, with the cost of modernizing the eastern *Länder* not covered by commercial investment having to be financed by further public borrowing. Governments everywhere appear to be in deficit.

But will not those problems melt away when economic activity picks up elsewhere, as it has recently shown signs of doing in both France and Japan? Then, the conventional wisdom has it, the demand for internationally traded goods will revive so that trade balances correct themselves, while more people will find jobs to do so that governments' tax revenues will increase and fiscal deficits will disappear at the same time. Would that it were that simple!

One of the rules of the game that have changed, probably for ever, is the scale on which funds are now invested across national frontiers. In the old days, when a successful business decided what to do with its surplus funds, it would usually invest them in its existing operation, making it larger or otherwise more productive. Now it is often easier, and potentially more profitable, to extend elsewhere or to buy a related business somewhere else (as the German car maker BMW bought British Rover Cars the other day). Emergent economies such as China are a powerful attractor for such

funds. So, on a smaller scale, are the countries of Central Europe and Central and Latin America. The total amount of direct investment in emergent economies may now be comparable with the US federal deficit, say \$250 million-plus a year. In the old days, of course, the same funds would have stayed where they were generated, improving the productive capacity of those who earned them. That is part of the real reason why the recovery from the recession is so slow.

But these are not the only overseas investments the prosperous world should now be making. Continuing confusion over the future of Russia is a reminder that the industrialized West never made good its promises to provide financial assistance for the reconstruction of the Russian economy. Instead, there is an emerging argument about the reasons for the growth of reactionary forces there. It is even more urgent that the West should take a decisive lead in reinvigorating the developing world, but that cause is also on the back burner, pending the awaited return of prosperity.

But what if the rules have really changed, and if prosperity 1980s-style does not float back with the return of spring in the Northern Hemisphere? The best bet is that private capital (mostly owned by the West's contributors to pension schemes and other savings schemes) will continue to follow commercial opportunities, while governments' obligations to others less fortunate will take up a substantial part of whatever extra wealth is generated. The view that that is what should happen cannot easily be denied, given the political dangers of continued neglect. That is why this recession will not end with a joyous bang, but with a long drawn-out whimper. □

Small country research

The European Union should pay attention to the research needs of smaller countries.

QUEBEC and Wales (part of the United Kingdom) have more in common than a language of their own. Quebec, with fewer than 7 million people, is a striking example of a small community seeking to make its way in the world by research; its provincial government has managed to find the resources usefully to complement those available federally, but there is some way to go (see page 389). Wales, with 4 million people, is a less autonomous part of the United Kingdom, but appears to share Quebec's belief of a quarter of a century ago that it is neglected by the central government in research



support. Whether or not the complaint is justified (see page 383), there is no question that Wales is one of several parts of the European Union in which more people (for their own interests and in the general good) should be engaged in research. Greece, Ireland and Portugal are others.

Quebec's remedy cannot apply in Wales or Northern Ireland; Quebec raises provincial taxes of its own, using them in part to support supplementary research funds, but Wales and Northern Ireland depend for all their revenue on funds transferred from Westminster to government departments based at the provincial capitals, Cardiff and Belfast respectively. Welsh nationalists are now demanding that 5 per cent of British research council funds should also be transferred to Cardiff for the Welsh Office to spend as it chooses, but that is quite misguided. Quite apart from the recently acquired reputation of Welsh public institutions for mispending public money, that would be a recipe for supporting research that would not qualify competitively for support. What Wales needs is a means for helping potentially able research groups to get started on research. (Northern Ireland has such a fund, perhaps only temporarily.) That is what Plaid Cymru, the nationalist party, should be pushing for.

In Britain as a whole, one of the defects of the present arrangements for research support is that it has been made constitutionally insensitive to regional considerations. Research grants are awarded competitively, while substantial contributions to overhead costs are channelled through the higher education funding councils on the basis of the regular research 'assessments' of university departments. Left to itself, this system can only further increase the concentration of research on well-established centres, exacerbating the sense of neglect at peripheral institutions and making it a self-justifying prophecy that their research-grant applications will always be uncompetitive. The remedy for Wales, as well as for institutions in England not in the mainstream of research, is that there should be funds to enable promising research groups to prepare to compete effectively, which is not a trivial undertaking. The goal is at once to increase the value and maintain the quality of research, not to foster the first at the expense of the second.

Some version of this recipe could be made to work on a European scale, but with the European Commission playing the central role. The obvious difficulty is that even if the commission were to offer initial support to promising research groups chosen transparently, by independent expert committees, there might be no research funds for which to compete. That is why, on the European scale, it would be preferable that the commission should follow the practice in the United States whereby central granting agencies make deals with recipient governments to share the costs of helping research get started. That way, it must be hoped, recipient governments would be shamed into turning promising ideas into well-founded research programmes. Without some such device, the regional disparities of which Plaid Cymru complains in Britain will be perpetuated on a much larger scale in Europe as a whole. □

Assassination in Mexico

Last week's outrage will at least give the United States a beneficent interest south of its border.

ONE thing is certain about last week's assassination in Mexico of the government party's candidate for president in the elections due next August, Mr Luis Donaldo Colosio: the US Congress would not have passed the North American Free Trade Agreement (NAFTA) if the assassination had already happened. Opposition to NAFTA was already strong enough to require President Bill Clinton to deploy the full battery of sticks and carrots with which US presidents are endowed. He would have found the task impossible in the light of such apparently glaring evidence of lawlessness (the attempt on the life of his own predecessor in 1981 notwithstanding). In the event, Clinton has followed the honourable and neighbourly course, offering to swap US\$6 billion for pesos if Mexico should need help on the foreign exchanges.

Thus is the course of history shaped. Six months ago, the assassination would have made a gulf between the United States and Mexico. Now, it has the opposite effect of drawing them together. Although most of last year's opposition to NAFTA in the US Congress stemmed from anxiety that jobs would migrate south of the border, Mexico's political system also attracted substantial complaint. To people in the United States, it is more than a mere cause for wonder that the same political party (the Institutional Revolutionary Party, or PRI) should have been in power for 65 years. President Manuel Salinas has done something to meet complaints from the north, but not so much as to risk his party's loss of the election. But NAFTA is an important feather in his cap. The best hope, for Mexico and for the rest of North America, is that his successor will have a chance to build a more thorough reform on that.

There is much to do. Mexico is an important and populous country of more than 80 million energetic people in transition from backwardness to modernity. Much of Mexico is rich, but there is too much obvious grinding poverty. NAFTA could generate the resources with which to complete the transition to a modern state as well as to make a lasting settlement with the Zapatista rebels in the Chiapas region of the south. (Already, the government's annual budget deficit of US\$20 billion a year is offset by the inward flow of investment funds.) The result could be not only a modern but a democratic Mexico.

The case for NAFTA has always been mostly economic. Not only Mexico but also its trading partners in the United States and Canada will benefit. But a side-effect, usually beneficial, of free trade is political and cultural interdependence. Last week's assassination will no doubt make the United States pay more, and more intelligent attention, to what is happening in Mexico and even further south. That, sad consequence of a man's needless death though it may be, will be in everybody's interest, not just Mexico's. Maybe we shall now see the beginning of the Pan-American policy people have been talking of for a century. □

Clinton courts research hospitals

Washington. President Bill Clinton is seeking support for his health-care reform legislation from the leaders of US academic medicine — a group not previously courted in what is going to be a bruising congressional fight.

During a private meeting with the leaders of Boston's many top research and teaching hospitals, the president promised that his bill will provide billions of dollars to subsidize the higher-than-average cost of care at institutions such as the Harvard-affiliated Massachusetts General Hospital (MGH), where the cost of taking care of patients is intermingled with the costs of doing research and training the next generation of scientists and physicians.

Although the president made no binding promise in dollars, Philip R. Lee, the assistant secretary for health in the Department of Health and Human Services, has estimated that \$10 billion is the least needed annually, while agreeing with medical school administrators that something like \$18 billion is a more realistic figure.

Health-care reform is built on the premise that the costs of seeing a physician or staying in a hospital can be driven down through

efficiency, cost controls and a form of competition that will induce people to flock to low-cost consortia of doctors and hospitals functioning on the theory that high volume can lower per unit (or per patient) costs. If



Clinton promises money to research hospitals — among many others.

such a system were to become universal in the United States, the country's research hospitals could be driven out of business.

During a trip to Boston, Clinton had a 45-minute meeting with a dozen medical policy

luminaries, including Mitchell Rabkin, head of Harvard's Beth Israel Hospital, Samuel O. Thier, president-designate of the MGH, Daniel C. Tosteson, dean of Harvard, and Kenneth I. Shine, president of the Institute of Medicine in Washington, DC.

According to those present, the president was at his political best in this small group. Well-briefed, he knew who had recently written about health reform, who had cancelled a meeting to come to see him, who was on his side and who was wavering.

Clinton made a clear commitment to federal support of the education and innovation that are the hallmark of academic teaching hospitals, quite apart from the funds they receive from the National Institutes of Health and other federal granting agencies.

For years, US teaching hospitals have lived off a complicated structure of cross-subsidies. For instance, Medicaid (the federal health insurance programme for the poor) has given hospitals a fixed per-patient premium or bonus to cover the cost of educating physicians-in-training. In addition, it has been common practice for hospitals to charge insured patients a kind of premium to make up for the cost of services to patients who cannot afford to pay.

These sources of revenue have been the life-blood of teaching hospitals, which treat more uninsured patients than do smaller, community hospitals, and which also treat the most seriously ill among the population. That is what their high-technology resources are for, but it also increases the bill.

Under Clinton's plan for universal coverage and a broad network of regional alliances — which are likely to die before legislation passes Congress (see *Nature* 368, 178; 1994) — the old pattern of cross-subsidization would disappear. One viable alternative is outright federal support. And this is what Clinton promised, urged on by Senator Edward M. Kennedy (Democrat, Massachusetts), chairman of the Labor and Human Resources Committee that will play an important role in fashioning health-care legislation.

The president's trump card in Boston was that none of the many competing health-care reform bills before Congress even mention academic health centres. Therefore, he said, academic medicine should support his bill. Speaking on behalf of the group, Rabkin says: "The consensus is that the president's bill is a winner."

But the president has made more promises lately about health reform to more groups than he can possibly keep, and bills competing with Clinton's are likely to include the care and nurture of academic medicine before the day is done. Thus, Clinton's promise is just a blip in the present fast-paced political debate. Nonetheless, it was a positive blip.

Barbara J. Culliton

Drugs bill to rein in profits

Washington. A bill introduced in Congress would amend existing laws to curb excessive drug company profits from monopoly supply of so-called 'orphan drugs' developed to treat rare diseases.

Sponsors of the bill say that it would still allow sufficient incentive to stimulate research on true 'orphans' — drugs of limited commercial potential aimed at treating rare diseases and disorders with a target population of fewer than 200,000.

The proposed legislation has bipartisan backing and was introduced in both houses last week by Congressman Henry Waxman (Democrat, California) and Senator Nancy Kassebaum (Republican, Kansas). Co-sponsors include the Democrats Ted Kennedy (Massachusetts) and Howard Metzenbaum (Ohio) in the Senate and Gerry Studds (Massachusetts) in the House.

Support for the proposed changes to the 1983 Orphan Drug Act has come from both industry and patient advocacy groups.

Under current law, an unlimited number of companies can apply for an orphan drug designation from the US Food and Drug Administration (FDA). But, in a first-past-the-post system, the first orphan drug to receive FDA approval is granted seven years of market exclusivity, during which time no other company can be licensed to market drugs to treat the same rare disease. A new provision in the act would shave three years

off the period of exclusivity, although this could be extended to seven years for drugs of very limited commercial potential.

Past attempts to amend the act have met with opposition from some segments of the drug industry. Efforts to introduce sales limits and to apply any amendments retroactively to drugs already enjoying market exclusivity were strongly resisted. The new provisions would apply only to drugs not yet in human clinical trials or on the market. Orphan drug status would be rescinded, however, if at any time the target population rose above 200,000. Market exclusivity would also be shared if two or more companies could demonstrate that they had developed their drugs simultaneously.

The original act was designed to provide special market and tax incentives to companies developing drugs for rare diseases. Few would dispute that the act has been a success: Waxman says that more than ten times as many orphan drugs have been developed in the 10 years since its enactment than in the previous decade.

Some of the earliest drugs to be designated as "orphans" — recombinant human growth hormone used to treat pituitary dwarfism, the anti-anaemia drug, Epogen, and Ceredase, developed to treat Gaucher's disease — turned out to be extremely lucrative for the manufacturers, leading to calls for new legislation.

Diane Gershon

Kessler adds heat to smoking debate

Washington. The head of the Food and Drug Administration (FDA), David Kessler, has told a congressional hearing that the tobacco industry may intentionally control the levels of nicotine in its products in order to create and sustain addiction in a majority of smokers. He also revealed that at least one tobacco company sponsored animal research in the 1980s that showed that nicotine was addictive. When the company learned that a

DOONESBURY

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

journal had accepted a paper on the topic from its scientists, it took out an injunction to prevent publication, he said. Kessler would not name the company.

The revelations came during a hearing before the health and environment subcommittee of the House of Representatives Committee on Energy and Commerce. Congressmen concentrated their fire on the key question of the manufacturers' intent, which is central to a fierce debate in the United States about whether the FDA can regulate tobacco products as drugs.

If there is an intention to create and meet an addiction, cigarettes can be regulated under the Food, Drug and Cosmetics Act of 1938, on the grounds that they are intended to alter the structure or function of the body. As a result the tobacco industry would have to withdraw its products — cigarettes and smokeless tobacco — and submit them to FDA efficacy and safety tests before being allowed to market them again.

If, on the other hand, nicotine is in cigarettes simply because it is a natural constituent of tobacco, tobacco products would not be subject to regulation by the FDA.

The debate sprang to life last month with Kessler's response to a petition filed by anti-smoking groups (see *Nature* 368, 83; 1994) arguing that by marketing low-tar and low-nicotine cigarettes, the tobacco industries were developing products intended to mitigate disease and that nicotine changes the structure and function of the body. Thus, cigarettes should be classed as a drug.

Kessler startled smoking and anti-smoking groups by saying that the petition might be right. Kessler cited the mounting evidence that nicotine is a powerfully addictive drug as well as evidence that the tobacco industry controls the levels of nicotine that satisfy this addiction.

In testimony last week, Kessler said that

"numerous patents illustrate how the industry has been working to sustain the psychoactive effects of nicotine in cigarettes". But Alexander Spear, vice chairman of the Lorillard Tobacco Company, told the hearing that holding a patent does not mean that the invention is in use.

Major cigarette manufacturers declined to testify, but Charles Whitley, a consultant for the Tobacco Institute, the industry's

trade association, told the hearing that the tobacco companies do not manipulate nicotine levels in different types of tobacco leaf. He said that although nicotine is extracted during manufacture, no more is added to the actual product than is present originally. The Philip Morris Company made the point even more strongly on the day before the hearing by filing a \$10 billion lawsuit against ABC television, which alleged that the company spiked cigarettes with nicotine to keep people addicted.

Kessler admitted that the FDA does not yet have sufficient evidence to establish the motives of cigarette manufacturers. But the anti-smoking groups are optimistic that the attention generated both by his remarks and by congressional hearings will help to achieve their main aim — tough federal legislation restricting the use of cigarettes and reducing the health damage they cause.

Helen Gavaghan

Quick solution demanded for Rutherford

London. The British government is being urged to make a quick decision on the future of the two main central research laboratories of the reorganized Science and Engineering Research Council (SERC) — the Rutherford Appleton Laboratory and the Daresbury Laboratory — in order to boost declining morale among staff.

Both the Royal Academy of Engineering and the Institute of Physics, in separate submissions to the government on the reorganization of the laboratories, express strong opposition to either laboratory being sold off to the private sector. Both suggest that they should be set up as a single 'non-departmental public body' (NDPB) under the Office of Science and Technology.

But each of the two bodies also takes the government to task for delays in implementing such a change, which was first proposed last summer by Sir David Phillips, then chairman of the Advisory Board for the Research Councils.

Clive Foxell, president of the Institute of Physics, says in a letter to William Waldegrave, the cabinet minister responsible for science, that it is "inexcusable" that no decision has yet been reached by the government. "Morale has suffered," says Foxell. "Too much of the laboratory's management time has been expended on dealing with this issue rather than on future activities within the laboratories."

The future of the Rutherford Appleton and Daresbury Laboratories has been up in the air ever since last year's government white paper (policy document) which led to the break up of the SERC this week (1 April). Early agreement was reached on combining the two laboratories into a single

administrative unit operating under a single director, but the government has not yet decided how far they should be exposed to market-place competition.

Earlier this year, a report from management consultants KPMG gave the laboratories high marks for the value of the services that they provide to universities throughout Britain and abroad, and concluded that privatization "would not be a sensible option".

Both the Royal Academy of Engineering and the Institute of Physics support this view. "We must ensure that unnecessary change does not take place for the sake of it," Sir William Barlow, the president of the academy, said last week.

One alternative to setting up the combined laboratories as an NDPB would be to hand over management to a consortium of universities, as is common practice in the United States. But the academy considers it "questionable whether UK universities have relevant experience or even the decision-making mechanisms to set up such a system".

Both institutions say that the two laboratories should be encouraged to increase the amount of contract research carried out for private industry, and that their role is likely to increase in importance as university departments seek the economies of scale offered by centralized research facilities. But both also support the government's desire to encourage greater competition between the providers of publicly funded research services (while warning against the application of this principle to basic research).

In a separate comment to Waldegrave, the Royal Society also opposes privatization.

David Dickson

Japan's *E. coli* genome project falling short

Tokyo. A pioneering effort launched by Japan in 1989 to sequence the complete genome of the bacterium *Escherichia coli* is still far short of its goal, and now lags behind two more recent international efforts, led by European scientists, to sequence the genomes of the bacterium *Bacillus subtilis* and yeast.

At the start of the *E. coli* project, its Japanese leaders said they hoped to complete the identification of the 4,700 kilobases of the genome in about five years (see *Nature* 338, 283; 1989). But an international symposium on large-scale DNA sequencing held in Tokyo earlier this month was told that only 274 kilobases have so far been sequenced. Another 500 kilobases have been sequenced by a US group at the University of Wisconsin, while many other *E. coli* sequences are scattered in databases around the world.

In contrast, the *B. subtilis* project, initiated by a consortium of five European laboratories in September 1989 and partially backed by the European Commission in Brussels, has sequenced about 900 kilobases, including 500 kilobases by seven Japanese laboratories.

Similarly, a recent worldwide effort led by Europe to sequence the yeast genome, and involving scientists in Europe, the United States, Japan and Canada, has already completed more than 2,000 kilobases of four yeast chromosomes (III, XI, II and VI). The symposium was also told that the whole yeast genome of 12,800 kilobases may be completed by 1997.

The Japanese *E. coli* project was given a special "priority" grant of several million dollars over a period of three years by the Ministry of Education, Science and Culture (MESC). But at the time it was launched, many scientists were still "cold" to the idea of systematic sequencing of genomes, says Katsumi Isono of Kobe University, one of the project's leaders.

As a result, says Isono, Japan was unable to win international participation. "Perhaps we did not push rigorously enough," he admits.

Another reason sequencing projects attracted little support among young researchers in Japan was that they feared they would be required to carry out the work of technicians. Furthermore, although about 30 Japanese researchers participated in the project during the three years of the priority grant, many did not do sequencing.

European scientists at the symposium offered two explanations for not participating in the Japanese project. One was that two US groups announced in 1987 that they intended to sequence the whole *E. coli* genome in two years, a move that deterred the formation of an *E. coli* project in Europe.

In addition, many European scientists felt that, given the number of individuals

working on *E. coli* worldwide, and the fact that many of its genes had been sequenced, a special project was not needed to complete the sequencing.

Another factor hampering the Japanese project was the rigid set of rules governing the use of grants awarded by the ministry of education. Unlike US and European government grants, Japanese government grants cannot be used to employ full-time personnel, such as technicians and postdoctoral fellows. As a result, the Japanese group was unable to train and then keep people with the

technical skills needed to carry out a long-term sequencing project.

After the priority grant ended in 1992, the project struggled along with a small grant awarded to a researcher at Tokyo University, which ends this month. Despite the setbacks, the project leaders are still hoping to win another "priority" grant next year, and they plan to subcontract much of the sequencing work to private companies. They also still hope that European scientists will join the project.

David Swinbanks

Regional Cinderellas ask for more

London. Figures published by the British government last week seem likely to reignite a long-standing campaign by scientists in Wales and Northern Ireland for more money to be spent on scientific research in outlying regions of the United Kingdom.

The figures for the regional distribution of funds spent by the five British research councils reveal that whereas Wales and Northern Ireland have respectively 5.0 and 3.0 per cent of the British population, they account for only 3.7 and 0.6 per cent of the total research council expenditure.

For some research councils, the difference is even sharper (see figure). The Sci-

ence and Engineering Research Council, for example, spends £8.3 per head of population in England (and £8.6 in Scotland), but only £3.75 in Wales and even less in Northern Ireland.

Research council officials explain the disparities by saying that the distribution of research grants is determined by scientific merit and by historical factors that have led to the creation of facilities at or close to major research universities in England.

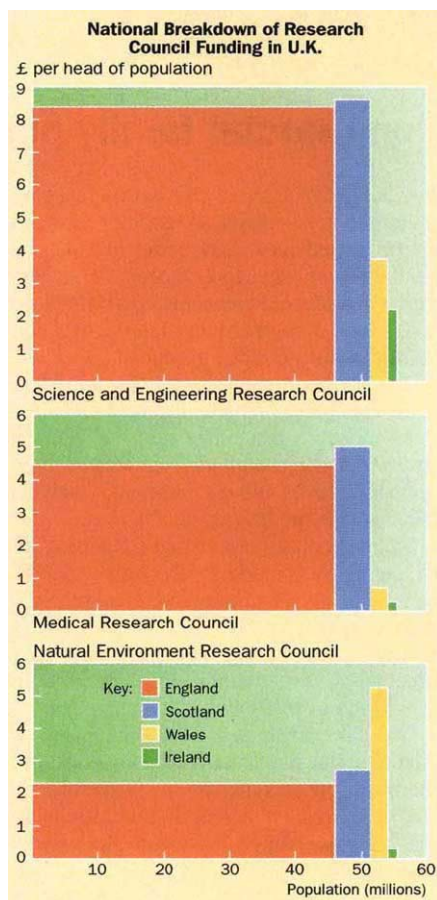
But some researchers complain that present policies underplay the value of government-funded research institutions in stimulating economic development, a factor widely recognized elsewhere in Europe and in the United States. "As far as I can tell, economic and regional factors have never been taken seriously in deciding where to establish major research institutions," says Phil Williams, professor of physics at the University of Wales at Aberystwyth.

Williams, a plasma physicist, says that countries as different in their political traditions as Sweden and the United States have located major astronomy facilities in depressed mining areas (Koruna and Greenbank, Virginia, respectively). He claims the British government has never acted similarly.

The regional distribution of British funds was made public last week by David Davies of the Office of Science and Technology in response to a parliamentary question from Cynog Dafis, Plaid Cymru member of parliament for the Welsh constituency of Ceredigion & Pembroke North.

The figures have been seized upon by Welsh nationalists, who say that research council spending should be integrated into regional policy. This pressure is likely to figure prominently in nationalists' campaigns in the elections to the European Parliament this June. Dafydd Williams of Plaid Cymru's political liaison office in London says that past practice on the siting of research establishments amounts to imposing "a reverse regional policy by default".

David Dickson



Monsanto sues over BST

Washington. Monsanto Company has sued two dairy producers for allegedly misleading the public by advertising their milk products as not coming from cows treated with recombinant bovine somatotropin (rBST).

The company, based in St Louis, Missouri, says it acted because the two companies failed to comply with the interim guidelines on voluntary labelling of rBST-free milk and milk products, issued last month by the US Food and Drug Administration (FDA) (see *Nature* 367, 585; 1994), which warn against the use of statements such as "from cows not treated with rBST" unless the claim is put in a proper context.

While Monsanto is asking the FDA and state regulatory authorities to enforce the guidelines, at least one state — Vermont — seems likely to respond to consumer concern by passing a law that will require special labelling for milk or milk products from rBST-treated cows.

Monsanto has filed suit against Swiss Valley Farms Company of Davenport, Iowa, and the Pure Milk and Ice Cream Company of Waco, Texas. In a written statement, Monsanto says that their rBST-free claims are potentially misleading and fail to tell consumers that there is no significant compositional difference between milk from treated or untreated cows. In short, "they falsely imply... that labelled products are safer, more wholesome, and in other ways superior" to milk from rBST-treated cows.

Monsanto is the only company with FDA approval to sell rBST, a bioengineered growth hormone that boosts cows' production of milk. The company sells it under the brand name Posilac. Opponents of rBST argue that consumers have the right to know if it has been used, and that dairies should therefore have the right to label their products as rBST-free.

Andrew Kimbrell of the Foundation on Economic Trends (FET), a Washington lobby group opposed to genetic engineering, says that by taking such a hard line with these two companies, Monsanto is hoping to snatch a quick victory to use as a precedent against the larger companies. It is "a classic bullying tactic", he says.

Even with interim labelling guidelines in place, some states may opt for mandatory labelling in the face of mounting consumer pressure. The state legislature in Vermont last week voted in favour of a bill that would require dairy farmers who use rBST to notify dairy processors and cooperatives of the fact, and would require mandatory labelling of all milk and milk products from rBST-treated cows sold within the state.

The FET has filed a legal petition with the FDA citing a possible conflict of interest with the agency's deputy commissioner for policy, Michael Taylor. From 1984 to 1991,



Taylor was a partner in the law firm of King & Spalding in Washington, DC, with clients that included Monsanto. The petition calls for an investigation of Taylor's work at the FDA by the inspector-general of the Department of Health and Human Services and by Congress, and requests that the agency rescind its interim guidelines on voluntary labelling, which Taylor approved.

FDA commissioner David Kessler sprang to Taylor's defence, saying that he broke no laws or regulations and has adhered to the right ethical standards at all times. Kessler says that for his first year at the FDA, Taylor agreed not to participate in any specific matters relating to Monsanto, and was not involved in the agency's decision to approve Posilac last November. **Diane Gershon**

High hopes for new Australian minister

Sydney. Australian scientists are hoping that the status of science policy in top-level government discussions will have been raised as a result of a ministerial reshuffle last week by Paul Keating, the prime minister.

The reshuffle was caused by the departure of several key ministers and means that Chris Schacht, previously minister for science and small business, will lose the science part of his portfolio to Peter Cook, the minister of industry. Cook is a senior and influential politician who (unlike Schacht) is a member of the 17-strong inner cabinet which makes all major government decisions.

Nothing has been publicly said about the reasons for the change in minister. But Schacht's two major proposals for reorganizing government science had created considerable ill-will among the scientists who would have been affected, and were blocked last year.

One of these was to create a Marine Institute out of various parts of the Commonwealth Scientific and Industrial Research Organization (see *Nature* 366, 97; 1993). This proposal is to be referred to a committee; but having lost a major supporter, its future is now in doubt.

As the head of an 'outer' ministry, Schacht had only limited access both to the cabinet and to Keating. He remains minister for small business and customs, and has been given additional responsibility for construction. **Mark Lawson**

Commercial family planning for India

New Delhi. Frustrated by the delays faced by government scientists in getting research results turned into marketable products, a group of reproductive biologists at the Indian Institute of Science (IIS) in Bangalore have set up a company to manufacture and distribute birth control products.

The new company, called Reproductive Biotechnologies Private Limited (RBPL), seeks to make the entire range of family planning and reproductive services — now monopolized by the government — available through the private sector.

The scientists responsible for setting up the company include P. R. Adiga, M. R. Moudgal and G. Padmamabhan, all full professors at IIS, and T. C. Anandkumar, a former IIS research scientist and past director of the Institute for Research in Reproduction in Bombay.

The creation of the RBPL, says Anandkumar, is in line with both the current liberalization policies of the government, and pressures on research institutions to raise money through links with the private sector.

Although scientists at IIS will act as

advisers to the company and take part in collaborative research projects, they will not be allowed to take up salaried posts with the company while they are still in the service of IIS.

According to Anandkumar, the company is also being supported by scientists working elsewhere in India on research into reproduction. Money for the venture will be raised through the sale of equity shares and commercial loans.

Anandkumar says that one reason for setting up the company was the difficulty faced by government-supported scientists in transferring their discoveries quickly to the marketplace. Both Adiga and Moudgal, for example, had developed a birth control vaccine more than a decade ago, and both will retire from IIS before their products reach the market.

RBPL intends to identify promising reproductive technologies that need a capital investment to be developed into marketable products. One such product which will be manufactured soon is a hormonal nasal spray, an alternative to the oral pill.

K. S. Jayaraman

Joint research projects 'need more regulation'

Paris. Pressure is growing in Brussels for Europe to develop a formal policy on regulating technological collaboration with the United States and Japan to ensure that, as such collaboration expands, Europe's commercial interests receive better protection than they do at present.

The move is being led by Anne-Marie Goedmakers, a Dutch member of the European Parliament who sits on its committees for budget and for energy, research and technology. "Europe tends to be naive about collaboration," she says. "We presume it's a good thing, and only think about the economic consequences afterwards."

Last week, representatives of the European Commission and European Parliament met in Brussels to discuss a report on cooperation with the United States and Japan. This was commissioned by the parliament's office of Scientific and Technological Options Assessment (STOA) from the Policy Research in Engineering Science and Technology unit (PREST) at the University of Manchester in the United Kingdom.

There was general agreement at the meeting that the European Union (EU) should not adopt a 'fortress Europe' policy. Indeed, one of Goedmakers' recommendations is that each committee managing the EU's major technology programmes — for example, in biotechnology, information technology, and transport — should be required to suggest possibilities for collaboration in its area, identifying both the benefits and costs involved.

The PREST report says that motives for collaboration include a desire for access to knowhow, the need to share the costs of expensive projects and greater efficiency in developing international standards or tackling global problems.

But it points out that there are also obstacles to global collaboration. These include concern about threats to competitiveness, a lack of consensus on the terms of sharing intellectual property rights (IPR) — especially between the United States and Europe — and an "institutional mismatch" between cooperating partners.

Intellectual property is, in particular, a thorny issue. Goedmakers says there should be international agreement on how IPR should be allocated within collaborative projects. This should always be defined from the outset, she says, as scientists involved in such projects often consider only how patents should be shared after the work is done.

Goedmakers also wants the EU to put "political pressure" on its partners to achieve better reciprocity in international cooperation. For example, she says it is unfair that European subsidiaries of US and Japanese companies can take part in EU research programmes, whereas, she claims, subsidi-

aries of European companies are excluded from some national programmes in the United States.

One conclusion to emerge from the meeting was that Europe should increase the number of fellowships awarded to industrial researchers to visit the United States and Japan. In particular, these would enable more industrial scientists in Europe to become familiar with other ways of managing research.

Christine Asmussen of STOA says there is concern in the European Parliament that the United States and Japan appear to make better use of the results of joint projects than European countries. "We need to get better at using the results of cooperation," she says. "If we can't get anything out of it, then

we shouldn't cooperate."

One difficulty facing any attempt to regulate collaborative projects is that EU-financed research projects account for only a few per cent of all spending on research in Europe, and most bilateral collaboration occurs outside EU programmes.

As a result, the meeting recommended that cooperation in research in general should be more closely monitored. "There is much collaboration already taking place," says one commission official. "It would be better to have it organized." He adds that giving the commission responsibility for doing this would tie in closely its powers under the Maastricht agreement to coordinate the research activities of the member states.

Declan Butler

Working scientists to get voice in Europe

Brussels. Antonio Ruberti, the European Union (EU)'s research commissioner, has unveiled plans for a new European Science and Technology Assembly (ESTA) to give working scientists a stronger voice in shaping the research policy.

Ruberti says the assembly — the concept for which has been floating around in various forms for more than a decade — will be in place by June. Although some critics fear that the body will be little more than an expensive talking shop, the few scientists already familiar with the plan are enthusiastic about its potential.

The assembly's main job will be to advise the commission on the direction and content of the EU's Framework research programmes, Ruberti says. But he also seems genuinely on the lookout for new ideas. At its own initiative, the assembly will be able to comment on any aspect of EU research. ESTA provides a "direct and permanent" link to the scientific community, he says.

The assembly will consist of 100 people — big enough to be representative, but small enough to be cheap and efficient, says Ruberti. To this end, the national research organizations will be represented through the European Science Foundation (ESF). ESF, which represents 55 research organizations, will nominate 24 scientists, from which the commission will choose 12. The commission says its role in the nomination process is designed to ensure that the assembly is distributed across countries and scientific disciplines.

The other members will be chosen from nominations from Academia Europaea (an organization of 1,500 individual scholars from 33 countries), the Association of National Academies of Europe (a loose federation of 71 academies, officially inaugurated last week in Paris), the standing conference of university rectors, and European research

organizations such as the European Laboratory for Particle Physics (CERN).

The assembly will also absorb the commission's existing advisory body, CODEST (Committee for the Development of Science and Technology), which has been criticized for letting its new role of managing a EU postdoctoral fellowship programme detract from its original aims, which are much like those of the new assembly.

A proportion of the seats in the assembly will be reserved for industrialists nominated by organizations such as the Industrial R&D Advisory Committee of the European Commission and the European Industrial Research Managers Association.

Peter Fricker, secretary general of ESF says he "welcomes" the assembly as "more representative". But he says the commission has not clarified what sort of decisions the assembly will take, nor what powers it will have. Fricker also understands that ESF's role in the assembly will not take away from its separate efforts to establish more formal links with the commission — it already has joint committees on programmes such as environmental change and ocean and polar research.

The member states, which have been unenthusiastic about past plans for an assembly, say they do not have enough details of the proposed assembly to comment. One UK official says that the main concern, both now and previously, is cost, and in particular fears that this would come from research budgets. He also adds that there is concern that the assembly could add another layer of decision-making and delay approval of programmes.

But Ruberti says that the members of the assembly will be unpaid and that the costs of meetings and other expenses will be met from the commission's general expenses.

Declan Butler

Slovak science lacks finance, direction . . .

Bratislava. Slovak science is fast losing its way amid the chaos that has reigned in Slovakia, the poorest of the four Visegrad states, since its split from the Czech Republic at the beginning of last year.

Prospects for the establishment of a proper governmental structure to support science and research, and the establishment of a legal basis for Slovakia's new grant agency, took another dip two weeks ago when the Slovak government fell for the third time since independence.

Science has had a hard time in the years following the collapse of communism in 1989 and the splitting up of Czechoslovakia at the end of 1992. The government unit for science and technology has been shifted between ministries several times (it is now hidden in the ministry of education, youth and sports), and has shrunk from 120 staff in 1988 to eighteen.

With no strong lobby to champion its cause in Slovakia's harsh economic environment, the total science budget has fallen from SK10 billion in 1989 to SK6.4 billion (around US\$150 million) in 1992, while annual inflation has averaged 20 per cent. The number of scientific staff has fallen from 63,000 in 1988 to around 30,000 today.

The country has no defined science policy, and Jaroslav Pokorný, currently head of the science section in the ministry of education, says that a new stronger government structure is necessary before he can consider setting science priorities for the country.

Despite the chaos above and around them, some scientists have made great efforts to improve the quality of their research. Most of the real reforms have taken place in the Slovakian Academy of Sciences, which now runs 57 basic research institutes. Unlike the situation in the universities, evaluation of its institutes — on the basis of objective bibliometric criteria which have resulted in withdrawal of state funding from eight — was started in 1990. A second round is to take place later this year. The pressure on individual scientists to improve their output has meant a significant increase in publications in the past three years.

The academy has reduced its total staff from around 6,200 in 1988 to 3,380 in 1993, partly as a result of brain-drain and partly through redundancies. But instead of being rewarded for its efforts, as it had expected, the academy has seen its funding fall at a faster rate than its staff. Its budget has nearly halved over the past two years — from SK668 million in 1992 to SK381 million this year.

The collapse in financial support, exaggerated by the high inflation rate, has put enormous pressure on academy scientists. Anna Pretová, director of the Institute of Plant Genetics in Nitra, can only hope that

her institute survives. Last year she managed to hold on to her 25 staff by cutting everyone's salary by 30 per cent. This year, even these measures were not enough. She had to fire five staff — "it was very difficult to do", she says — and numbers are now at the minimum allowed for an independent institute. She relies heavily on grants from the European Union and collaborative work with scientists in Western Europe to keep experiments ticking over.

One such Western scientist is Jan-Peter Nap, at the Centre for Plant Breeding and Reproduction Research in the Netherlands, who says that his continuing collaborative work on improving reliability of gene ex-

pression in transformed potatoes with the Nitra scientists has been rewarding. "Although they have virtually no facilities, they solve scientific problems with simplicity and elegance," he says.

The academy is pinning its hopes on the SK110 million that the government promised earlier this year to the new Slovak Grant Agency, set up in 1990 to fund research projects on a competitive basis, once it has achieved legal status. But the passage of the Grant Agency Bill through parliament will be delayed by the current political turmoil. In the meantime, no money is available to fund grant applications that have already been evaluated.

Alison Abbott

. . . while Czech budget doubles

Prague. The Czech government has agreed in principle to nearly double spending on science over the next few years. Combined with broad reforms that have recently taken place within the Czech Academy of Sciences, the move could give the country one of the strongest research bases in the former communist block.

In a roundtable meeting two weeks ago with scientific advisers and the finance and economics ministers, Czech prime minister Václav Klaus agreed that the country should raise its current level of spending on research and development from 0.45 per cent of gross national product (Kr4.1 billion, US\$140 million) to 0.7 per cent (predicted to be over Kr10 billion), over a period of around five years.

The decision is not yet official, but Igor Némec, minister without portfolio who is also chairman of the government science and technology advisory group, says that "the principles of the agreement will not change". This means that scientists may expect a rise of 20 per cent in real terms in next year's research and development budget. This should be confirmed by the end of June.

The Czech Republic has been relatively slow in comparison with other Central European countries to restructure its science. With a privatization programme now finally well under way, much of the applied science base of the Czech Republic has been lost. The total number of people working in scientific institutes, mostly in applied research, has fallen from 133,000 in 1989

to around 60,000, and the number is still falling.

But in the last year, the basic science research base, housed mainly in the Czech Academy of Sciences research institutes, has taken decisive steps towards rationalization. When the academy's new democratic council came into being in 1990, it proved to be ineffective in allowing true reform to take place, and eventually lost support even from within its own ranks.

When it was informed in autumn 1992 that its new budget was to be reduced by one third, however, things began to change rapidly. Within weeks, evaluation committees were set up and each institute was assessed according to its scientific output (number and quality of publications). As a result, more than 25 institutes were closed and staffing levels at the remaining 59 institutes halved.

Surviving scientists say that the swift and sweeping changes were not resented, but were generally felt to be long overdue. Their average number of publications has risen in the past year despite the reduction in support staff.

And the shedding of staff was not a major moral problem. With 0.45 per cent unemployment in Prague, nearly all dismissed staff found new jobs immediately: "and they are getting twice as well paid as us poor fools still doing science" jokes Jan Bureš of the academy's Institute of Physiology.

But academy president Rudolf Zahradník says that the reforms have still not gone far enough. "Although the evaluations done last year were valuable, and decently carried out," he says, "they were done in a hurry." He now wants to establish a series of evaluation boards with foreign scientists in the majority to look again at the efficiency of each institute, and would repeat the exercise regularly every few years.

Alison Abbott



Zahradník: reforms

AIDS plagued by journalists

■ This is a shortened version of a letter sent to *The Times* (London) but not published.

SIR — It was rather bewildering to read the article in *The Times* of 18 December 1993 by Simon Jenkins, that the main problem with AIDS is that it is “plagued by politics”. Jenkins fails to appreciate that AIDS is caused by a viral infection of unparalleled and horrendous consequences to the infected individual in particular and the human race in general.

In the case of other diseases caused by microorganisms that affected large numbers of people before the discovery of antibiotics and the development of protective vaccines, most of the affected died within a relatively short time after infection, whilst those who survived acquired — by and large — a lifelong protective immunity and were not infectious to others. In contrast, infection with the agent that leads to the development of AIDS invariably results in an incubation period of several years, during which time the individuals may be unaware of their infected and infectious state and infect their sexual partners.

Although HIV was first isolated in France 11 years ago, we have learnt an enormous amount about this virus and the way in which it causes a breakdown of the patient's immune system. The recognition of its importance, and corresponding financial support, led to the discovery that HIV infection invariably leads to AIDS. As a result we are also in a position to appreciate why, despite enormous efforts and 9 years of investment in vaccine development, we are no nearer to having either a protective vaccine or an effective drug that could prevent the progression to AIDS, let alone provide a cure.

The application of molecular biology to HIV research has enabled us to recognize that, unlike other infectious agents, the causative agent of AIDS continuously mutates in the body of the infected. This has led to the recognition that each individual HIV has its own molecular fingerprint. It also explains why, for example, AZT has, at best, short-term benefit, because the virus in AZT-treated individuals develops mutants within 6–12 months which become resistant.

It is bewildering that Jenkins can write that HIV infection does not lead to AIDS when we already know that the deadly disease develops in all heterosexual recipients of contaminated blood or blood products who do not belong to the *Sunday Times*’ “lifestyle high-risk group”. In the “high-risk group” we already know that more than 90 per cent will develop AIDS within 10 years of infection. Jenkins also fails to appreciate that more than 90 per cent of the men and women around the

world who have contracted AIDS do not belong to the *Sunday Times*’ “lifestyle” group because they contracted it during normal heterosexual intercourse. It is, therefore, incomprehensible that Jenkins can write that “no causal chain has been proved” between HIV and AIDS when it had been established some 10 years previously.

It is difficult to understand why Jenkins repeats the *Sunday Times* accusation that *Nature* has “led a propaganda war that may have unnecessarily stigmatized millions of HIV-positive people”. It is not unreasonable to expect the previous editor of *The Times* to know that progress in any discipline of science or medicine is based on earlier knowledge, and that any claim to discovery always has to be reproduced by others in the same field before it is accepted. For Jenkins to assume, therefore, that all the physicians and medical scientists involved in AIDS research and treatment are wrong is absurd.

Jenkins argues that we should take note of what Dr Peter Duesberg has to say, as if we were dealing with a philosophical or theological rather than a scientific and medical question on which, as said, all the necessary evidence has been provided to prove some 10 years ago that HIV causes AIDS. Jenkins also repeats the *Sunday Times*’ absurdity that HIV-infected haemophiliacs are dying because “they are bombarded with alien blood”, not realizing that haemophiliacs who are HIV-negative have a normal lifespan, while almost all those who become infected with HIV develop AIDS and die early, even children and teenagers who have been “bombarded” for a short period of time.

Another incomprehensible argument by Jenkins against the role of HIV in AIDS is, “why, after 10 years’ intensive research, are we no nearer the answer?” We have made considerable progress in understanding the disease, although little progress towards a cure — but the same is true after many more years of research into the common cold, various herpes infections, some of the major cancers and multiple sclerosis.

Jenkins quotes recent statistics from Cambridge about HIV infection but fails to note that, in the previous year, 37 per cent of new HIV infections were in heterosexuals while, early on, the UK infection rate in heterosexuals was in only single figures. The number of AIDS cases in Britain, 5,000, reflects the HIV infections that took place 7–10 years ago, not those occurring now, whose consequences will be seen between 2000 and 2005. Meanwhile, the present number of people with AIDS creates a false sense of security, aided by campaigns such as those of the

Sunday Times, encouraging people to disregard the risks and spread the infection further.

Largely through heterosexual intercourse, the spread of HIV infection and AIDS has already had a devastating effect in some parts of Africa, about which Jenkins remains silent. Heterosexual HIV and AIDS are also increasing at an alarming rate in countries such as India, Brazil and Thailand. The rates of HIV infection are lower in the West but are continuously increasing: in some US cities, AIDS is becoming the major cause of death even in women aged 25–40.

It is about time that *The Times* and the *Sunday Times* realized that, through badly informed and irresponsible journalism, they could mislead millions of their readers about HIV infection and AIDS. In the absence of a cure, and without a prospect for an effective vaccine, prevention is the only defence against this formidable disease. Informed, responsible journalists can play an important role in reducing the rate of spread of AIDS.

A. Karpas

Department of Haematology,
University of Cambridge,
Hills Road, Cambridge CB2 2QH, UK.

SIR — Because of your interest in our reports on HIV and AIDS, I should like to make it clear that, contrary to some reports, the *Sunday Times* has never presumed to pass judgement on whether or not HIV causes AIDS. We do however believe that a strong scientific challenge has been mounted to the consensus view that HIV is the cause of AIDS, and consider the issue so important as to merit wide discussion and examination. We will continue to report views and findings relevant to this debate.

Many such developments occurred over the past 12 months, and as we have been almost alone among the media in challenging the orthodox view that accepts HIV as the cause of AIDS, we have often found ourselves left with a clear field in writing about them. This has given the impression that we are running a campaign on the issue, but the view here is that all our articles have been of legitimate news interest, given our fundamental position that there are real uncertainties over AIDS causation.

I should like to add that my own doubts about HIV's role have grown stronger over the 20 months since we first set out the “dissenting” views, not least because of the unreasoning way mainstream science has reacted to the challenge, at first dismissive and then, when the “problem” refused to go away, openly censorious.

Neville Hodgkinson

(*Science Correspondent*)

The Sunday Times,
1 Pennington Street,
London E1 9XW, UK

Cancer institute Sex determination treated unfairly

SIR — Your report (*Nature* **368**, 89; 1994) of research-related budget allocation by the Higher Education Funding Council for England (HEFCE) for the years 1994–95 states that only two universities — Oxford and Cambridge — have been capped. The report is incomplete: the Institute of Cancer Research, a member of the British Postgraduate Medical Federation of the University of London, has also been capped, severely so.

The institute has been a higher education institution (HEI) for more than 40 years, but until recently was excluded from funding by HEFCE, or HEFCE's predecessors, for reasons that have been historical rather than rational. This anomaly was corrected in August 1993, when HEFCE funding of £2.4 million per annum commenced following ministerial transfer of exactly that amount from the budget of the Medical Research Council (MRC) to HEFCE. The sum involved was that contributed at the time by government (MRC) to the institute.

The MRC, despite this raid on its budget, treats us exactly as all other HEIs: grant applications are received and, when they are successful in competition against others, grants are awarded. Not so HEFCE. Despite our excellent research performance in the 1992 University Funding Council Research Assessment Exercise, our HEFCE grant is pegged at £2.4 million a year, with not a penny for PhD students, of whom we have 60.

Our total annual budget is £23 million, and our average competitively won grant income per academic staff member, of whom there are 80, is probably the highest in the United Kingdom. The imbalance between the two limbs of our dual support is as unsustainable as it is indefensible, yet HEFCE offers no reconsideration before completion of the next (and delayed) research assessment exercise in 1996–97. In the meantime, HEFCE continues to profess its commitment to competition and to selectivity based on excellence — and spends more than £500 million per annum on uncapped or even protected research support for HEIs that came below us in the 1992 assessment exercise. If HEFCE has a responsibility to use public funds for infrastructural support of the best available research in the higher education sector, then that responsibility is clearly capable of being better discharged.

Peter Garland

(Chief Executive)

Institute of Cancer Research:

Royal Cancer Hospital,

Fulham Road

London SW3 6JB, UK

SIR — A leading article (*Nature* **361**, 283; 1993) argued that a service for sex determination offered by a London clinic was so unreliable that it did not justify a debate on novel “serious ethical questions”. Since then, it has become possible to identify the sex of human pre-implantation embryos after *in vitro* fertilization (IVF). In our opinion, that is a serious development.

Although the use of IVF for nonmedical sex selection is held by some to respect a woman's autonomy, we hold that sex is not a disease and that sex selection is not a condition that justifies the use of invasive procedures. In our view, although woman's autonomy should be respected most of the time, clinicians, in accordance with reasonable professional standards, must refrain from actions that might impair rather than promote the interests of the patient.

The use of IVF for sex selection is a misuse of costly medical resources, at least so long as there are women who need this service, and who cannot afford it. But it is also troubling that the use of IVF for sex selection might establish a dangerous precedent.

That, in our opinion, is why IVF should be used for medical purposes only. It is not the idea of sex preselection in general that we oppose. If there were a simple noninvasive technique for sex selection not requiring the active intervention of medical staff and costly medical resources, then every couple would have the right of selection. But while sex selection is determined by IVF, the use of this technology by medical personnel for non-medical purposes raises serious ethical questions.

Asher Shushan

Josef G. Schenker

Department of Obstetrics and

Gynaecology,

Hadassah Medical Organization,

Kiryat Hadassah,

PO Box 12000,

Jerusalem 91120, Israel

Academic research a dirty word?

SIR — I was interested in Daedalus's suggestion¹ that gases might be stored in a molecular sponge. Our recent work on zeolites^{2,3} (not polymers in the sense implied by Daedalus) has produced strong evidence that some zeolites behave in a not dissimilar way under the stress of physical adsorption. In the physical adsorption of argon and nitrogen into silicalite, for example, the pores first of all fill to their maximum capacity and then the framework expands to accommodate

an extra 30 per cent of the adsorbate. This interpretation of experimental adsorption isotherms originated from simulation studies, and has been confirmed by very recent neutron scattering work (N. J. M. Tosi-Pellenq & J.-P. Coulomb, personal communication). Unfortunately, our proposal to extend this investigation by studying the conditions under which adsorbent frameworks might distort during adsorption did not receive support from the Science and Engineering Research Council. The referee commented that the work was “academic” and of no industrial significance. Perhaps the far-sighted research managers of DREADCO might care to contact me to discuss this project.

David Nicholson

Department of Chemistry,

Imperial College of Science,

Technology and Medicine,

London SW7 2AY, UK

1. Jones, D. *Nature* **367**, 518 (1994).

2. Pellenq, R. J.-M. thesis, Univ. London, 1994.

3. Pellenq, R. J.-M. & Nicholson, D. *Proc. COPS III Meeting*, Marseille, 1993 (in the press).

Freedom of proof

SIR — Vladimir Koliadin (*Nature* **367**, 406; 1994) offers a good insight on the danger of “democratic centralism” in the scientific establishment. However, he asserts that “(t)he best way to proceed is not to try to ‘prove’ or ‘disprove’ a theory, but to concentrate attention on the positive aspects of any theory”. I suggest, rather, that there should be no hindrance to complete freedom of any scientist to try to ‘prove’ or ‘disprove’ any theory, or to examine either ‘positive’ or ‘negative’ aspects of any theory. All opposition or hindrances to this total intellectual freedom of every individual scientist constitute ‘politically correct science’, which surely should be odious to every scientist.

Robert E. Kofahl

1322 E. Wilson Avenue,

Glendale, California 91206, USA

Early ‘cruise’

SIR — The Second World War German V-1 was not a rocket (*Nature* **367**, 424; 1994) but a jet-propelled pilotless aircraft. The next generation of terror weapons, the V-2, was an authentic rocket.

My recollection of the V-1 is vivid because one fell near me in England in 1944. The ‘buzz-bomb’ flew with a characteristic flutter until the engine shut off, and then it dived to Earth. (During this interval one held one's breath.)

Maxwell Gordon

Aji-Pharma USA, Inc.,

500 Frank W. Burr Boulevard,

Teaneck, New Jersey 07666, USA

Can Québec's science (and Canada's) survive constitutional separation?

The chance that *Québec* will separate from the rest of Canada grows as time passes without persuasive steps to keep the country together, but science is unlikely to be a casualty of separation.

Montréal. Just a few weeks ago, the snow was melting from the sidewalks of this majestic city, whose name is taken from the escarpment running parallel with the St Lawrence (itself a majestic flood) called *Mont Royale*. Now, with signs of spring, it will be open political season again. The spirits of those who long for separation from the rest of Canada, regarded as the honest alternative to the demand for "sovereignty" that has dominated the arguments of the past several years, will begin to rise again.

This year, like most years in the recent past, will be critical. Indeed, there is some urgency. The present provincial government of *Québec* must call an election before November and, on the calculation that delay is likely to exacerbate the constraints to which it is subjected, its opponents believe it will invite voters' opinions sooner rather than later. The present government of *Québec* is formed by the Liberal party (of which Pierre Trudeau, a *Québécois*, was perhaps its best-known federal prime minister). The present federal government of Canada is also Liberal, under prime minister Jean Chrétien; his name shows his provincial allegiance.

In *Québec*, the chief opposition party is the *Parti Québécois*, or *PQ*. Confusingly for outsiders, in the federal parliament in Ottawa there is also (since the decimation of the Conservative party in last November's federal elections) a powerful contingent of people labelled the *Bloc Québécois*. The bloc is there to influence events, the *PQ* to drive them. The *PQ*'s declared objective, if elected to form the provincial government, is to call for a referendum on separation within a year. At the beginning of March, the party's leader, Jacques Parizeau, startled even some of his supporters when he told an audience in *Montréal* that, if the second referendum did not yield a vote for separation (there was an earlier vote in 1980), his party would try again, and again, and . . .

It goes without saying that the question of whether or not to separate does not rest exclusively with *Québec*; unilateral secession led, after all, to the American Civil War. At the very least, the rest of Canada

would expect to have a say in matters such as the division of mutual assets (and obligations), which implies a protracted negotiation. The function of a referendum would be to express the will of the people of *Québec*, which is likely to be all the more influential if the federal government is itself weakened.

The case for separation, when put well, is to do with language, the French language in particular. *Québec* is an historical relic of the times in the eighteenth century when Britain occupied some of North America and France much of the rest. When Napoleon, in 1803, sold the territory west of the Mississippi to the stripling United States, the British and French were still fighting each other in Canada. France may have lost the war for Canada, but it left behind a booby-trap, which has now come alive.

cois had a rotten time for more than a century and a half. Until the 1960s, *Québec* was a relatively impoverished part of Canada, with little in the way of stylishness except *Montréal*. Part of the origin of this language war was the readiness of English-speaking Canada to dismiss as unimportant its curiously accented compatriots. For nearly half the *Québécois* now alive, direct experience of those times is within living memory.

Whence the pride with which *Québec* boasts of what has been accomplished in the past quarter of a century. What seems to have happened is that the provincial government has seized the constitutional opportunity to spend its own funds on self-improvement. That accounts for the wave of expansion in higher education beginning in the 1960s. But even before then, *Québec* had begun spending money

provincially on support for its own research. Then, as the risks of separation have increased, federal Canada has done what it could to be more than fair to *Québec*. As the years have passed, several important industries have sprung up in the province.

So what is the objective of separation? There is a preliminary point of some importance: separation is not a foregone conclusion. When asked directly, professional people tend to say they hope it does not happen. Some add that "it need not have come to this", explaining that everything would have been simpler if only the politicians engaged in the endless constitutional talks in recent decades had been readier to accept *Québec*'s

distinctive needs. (That is not as simple a demand as it may sound, especially when *Québécois* sometimes ask that the federal government, which operates one of the most liberal immigration policies in the world, should allow them to exclude those who cannot speak French.) Some even wish that, even at this late stage, a statesman of stature may emerge to discover a palatable solution — and then shake their heads over the way in which Canada's expectations of its politicians were destroyed last November.

Briefly, it is as if people have become tired of the argument, which has lasted a long time. In a sense, the objective of



Intelligent, educated and invariably bilingual people say that they feel differently in themselves when talking French. They do not have to protest that much. They slip into the *patois* easily to discuss what to tell a visitor, for example, or whether the duck is preferable to the chicken on the menu. *Francophonie* is not an operational name for the process of speaking French, but a synonym for 'kinship'. The epithet 'French Canadian' is no longer used, but the term *Québécois* does not with equal veracity apply to the French-speaking population of New Brunswick, a little nearer the Atlantic.

It goes without saying that the *Québé-*

separation is . . . separation. It is not so much that *Québécois* want nothing to do with people whose mother-tongue is English, or even that they wish to abandon the same language in professional discourse (which would be impossible), but that they want to live more intimately with people like themselves.

What would a separated *Québec* be like? Superficially, it need not be very different from the present. The provincial government, already strong, could easily take over the functions now managed from Ottawa. On matters dear to hearts in Ottawa, selfconscious concern for the environment and generosity on foreign aid, *Québec* would not change tack. Politicians and their supporters also insist that a separated *Québec* would continue, as in the past quarter of a century, to lavish funds on the education of the young and on technical modernity. Such doubts as there are centre on questions such as "Could we afford it?"

Upon separation, *Québec* would also have a foreign policy of sorts. On a visit to New York and Washington on 3 March,



the leader of the *Bloc Québécois*, Lucien Bouchard, declared that *Québec* would readily negotiate to become a separate member of the North American Free Trade Agreement and allow the United States to continue testing cruise missiles in the northern wastes of the province.

Whether separated *Québec* would burden itself with a defence policy is another matter, although there was much resentment earlier this month when the federal government, as an economy measure, made public a plan to close the province's only training college for officers in the armed forces: to some, it seemed a plot to undermine the defences of a separated *Québec* before it had come into being.

Two powerful impressions stand out from several conversations, mostly with academics and administrators. First, although professional people appear uniform in the view that separation would be regrettable, most of them volunteered that the issue is not for them to decide, but for "ordinary people". That was often offered as a reason for not being active in

politics. The irony is that the chief casualties of separation (if any) are likely to be blue-collar and unskilled workers, who would first feel the draught if separation meant fewer opportunities for jobs, or less investment in *Québécois* industry. The professionals seem confident that they will continue to be as mobile as they are now.

Second, a senior civil servant agnostic on the eventual outcome powerfully put the view that, whatever happens next in Canada, there must certainly be a profound change in the way things are run. The election in November was not so much a defeat for the Conservative federal government (although it was that) as for the pre-existing political system, which cannot be recreated. On that view, the question is not whether *Québec* can make its way in the world alone, but what will happen in any case to the rest of Canada. Those in Canada outside *Québec* who believe that separation would simply mean that they no longer have to be able to speak French are mistaken.

A further consideration is that *Québec*, whether independent or not, is in no sense a part of Europe, whatever its language. "We are part of North America", said

one, insisting that *Québécois* will continue to seek work and fortune where they can find them. They are inveterate entrepreneurs, whence their successes in the past few decades.

In science and technology, it is also clear that *Québec* will do everything it can to sustain its present interest in modernity. The achievements of recent decades, which are remarkable, will be continued if only the funds can somehow be kept flowing. The most obvious danger is that the consequences of serious misjudgement for a separate country of merely 6.5 million (roughly the population of Switzerland), will be irremediable. Small may be beautiful, but is not necessarily safe.

But the other side of that coin is the sheer enthusiasm for the idea of *Québec* that the past few years have inspired. There is an infectious gaiety about the place, while much of the rest of Canada remains glum. This stems from the remarkable sense of confidence the past few years have brought (but which may well be false) and from the growing sense that there will not be a Canadian solution on offer.

Can Québec stand alone?

Could *Québec* survive on its own? In strictly money terms, there may be no great difficulty. The provincial government boasts that its gross domestic product or GDP (C\$158 million in 1993) puts *Québec* between Turkey and Denmark in the value of its output, and between Norway and Austria in GDP per head of population. By the second yardstick, a separate *Québec* would be the eleventh richest country in the world.

There is also great delight in *Montréal* at the changing pattern of output in the province. Minerals and other mined products are a declining share of total output, although paper and paper pulp remain the most valuable single export, followed by refined aluminium. But the government delights in the increasing share of output derived from industries that add value to primary materials, notably in electronics and aircraft.

The claim that *Québec* can make its way in the world by exports of high-technology products is partly sustained by the circumstance that roughly half of the province's output is exported, either to the rest of Canada or elsewhere. Almost 70 per cent of *Québec's* manufactured exports end up in the United States.

But there are other *Québécois* phenomena that may profoundly augment future wealth. In the past few years, a number of *Québec* companies appear successfully to have turned themselves into multinational organizations. One of these is the transport company Bombardier,

which manufactures railway trains and locomotives in *Québec*, but which has recently made a business of buying up nearly failed companies elsewhere in the world and turning them around. (One beneficiary is the British aircraft manufacturer, Short Brothers or Belfast.)

Similarly, the company called Videotron, which supplies cable television services in Canada, is now busily digging up streets in London and elsewhere in Europe with the same goal — but with money borrowed locally. For *Québec*, the profits of these enterprises may count as 'invisible' exports. Certainly they have fired many in *Montréal* to regard free-wheeling international enterprise as their strong suit.

The provincial government, meanwhile, is more interested in what can be accomplished on the ground, in *Québec*. There are two pillars in its strategy for making *Québec* prosperous: to strengthen higher education and to encourage spending (and to spend itself) on research. In the second half of the 1980s, total research spending (by industry as well as government) increased faster in *Québec*, at 4.5 per cent a year, than in all other industrialized countries except Japan and Italy. (The rate of growth in France during the same period was also 4.5 per cent a year.) *Québec* is especially proud that industrial spending on research and development in the same period increased at an annual rate of 10 per cent (helped, no doubt, by a generous tax credit).

Thus, unlike most other governments, that of *Québec* is forever referring to the proportion of GDP spent on research and development, and not approvingly: the refrain is that the present percentage (1.68 per cent, compared with 1.49 per cent for Canada as a whole, but — a little shamefully — with 1.89 per cent in Ontario) must be increased.

How can that be done? In basic research, the strategy is that the research funds available competitively from the federal government, from the Canadian Medical Research Council for example, should be supplemented by funds available from *Québec's*, which at present amount to about 30 per cent of all receipts.

On development, *Québec's* action plan is more specific. The government has singled out a number of industrial sectors, called "clusters", for special attention, and (among other things) has busily set about creating research institutes, usually at universities, to carry out strategic research on their behalf. Biotechnology and electronics are among the clusters that have profited. Otherwise, the Ministry of Industry, Commerce and Technology is seeking to foster networks of interested companies in the belief that the outcome could be as fruitful a relationship between industry and government as the much-envied link between Japanese industry and the Ministry of International Trade and Industry in Tokyo.

The idea is simple, but the scheme has not yet been in being for long enough to tell how effectively it will function. For the industries concerned, the ministry and the groups of companies concerned set out to identify the industrial functions that are necessary parts of a free-standing integrated industry.

Aerospace is a good example, if only because *Québec* has 45 per cent of all Canada's business in the field (worth C\$3.1 billion a year). More than 70 per cent of the output is exported. Discussions between the government and industry have, for example, uncovered the link between aerospace manufacture and the general metal-forming engineering sector and have led to the conclusion that *Québec* is better able to supply aircraft manufacturers or satellite builders with cutting tools than to carry out metal-cutting work on a subcontract basis. The hope is that continuing discussion will stimulate the development of services in the second field.

Government officials acknowledge that this procedure is untried. To the extent that one of the goals is to create within *Québec* self-sufficient industrial groupings, the strategy may even be inherently uneconomic. Even if separation comes about at some stage in the future, there will presumably be no reason why manufacturers in *Québec* should not continue to buy in components and services from what

is now the rest of Canada, for example.

The reckoning is that *Québec* industry is already strong in the fields of aerospace, pharmaceuticals, information technology, heavy electrical engineering and metal and mineral processing. The criteria for these appellations is that there should be at least one company manufacturing to international standards and supported by an appropriate array of smaller companies. The cluster strategy can only assist them to be stronger.

The risks are likely to lie at the other end of the spectrum, among the less well-established industries, where government officials are likely to find themselves in the familiarly invidious position of "picking winners". Take textiles, for example. That *Québec* has a substantial interest in the field, with total employment in the province of 93,000 and products worth more than C\$8 billion a year, is not in dispute. The question, rather, is whether promised "government initiatives" can help to sustain the health of an industry whose success depends crucially on the success of specialist textile companies increasingly forced, as in most developed countries, to look for market niches in which they can survive.

Inevitably, for a government that is instinctively interventionist, grand strategy can from time to time be submerged by short-term considerations. In that spirit, the provincial government is just half-way through a C\$1 billion three-year programme intended to provide further stimulus to high-technology industry, and to create extra jobs in the process. This will no doubt not be the last occasion on which the government of *Québec* is diverted from long-term strategy by the need for politically palatable sticking plaster.

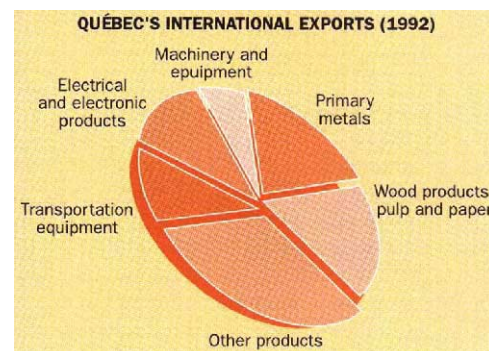
Even so, the provincial government's direct support for industrial research and development is only small, just over C\$10 million a year; the tax credit for research and development (at roughly 40 per cent of expenses) is worth more than C\$150 million a year to qualifying companies, some 10 per cent of the total value of research and development by business enterprises in 1990.

Over the past decade, the most spectacular growth has been in the pharmaceutical sector, in which the established multinationals such as Merck, DuPont and Glaxo have chosen *Québec* as the sites of subsidiary operations, offering access to both the Canadian and US markets; research and development spending in the province is now running at roughly C\$200 million a year, in part at least because of an undertaking by the drug companies to increase the volume of their research in Canada as part of a deal last decade with the federal government on drug prices. In the circumstances, the

annual value of the research and development may not be a sure guide to future volumes of production and of exports from *Québec* to the rest of the world.

In the circumstances, the crucial economic question whether a separated *Québec* would be viable must remain an open question. More exactly, since the provincial government's commitment to high technology will presumably persist even if *Québec* remains a part of Canada, the question is whether separation would decisively and favourably shift the balance.

One crucial factor is the degree to which a separate *Québec* would attract inward investments not otherwise likely to be made. Under NAFTA (the North American Free Trade Agreement), to which *Québec* supposes it would continue to belong, the province offers access to the US market, although not as economically as, for example, direct investment in Mexico. Even so, *Québec* may turn out to have special attractions for French companies



planning to establish subsidiaries in North America. Time (and separation) alone will tell how compelling is that attraction.

A second consideration is the extent to which separation would put further strains on what would then be the national budget of the new state. As things are, *Québec* is no more prosperous than Ontario, its neighbour to the west, nor is it more dependent on high technology for a living. On the assumption that separation would relieve *Québec* of its present share of the costs of supporting less prosperous parts of Canada (notably the maritime provinces) through the federal budget, the new state's budget would be affected chiefly by the degree to which that saving is offset by the extra costs of creating a new bureaucracy.

Little thought appears to have been given to these questions, but the biggest hazard for *Québec* might well be that a newly separated government would be tempted by the novelty of its situation to indulge in optimistic schemes for development whose effect was to run the public finances into further deficit than they are already, encouraging inflation, which is one of the chief enemies of industrial innovation. That, again, is a risk that will be assessed only if, and when, there is a separate *Québec*.

Diversity by episodic growth

Québec's commitment to higher education is most apparent in the crowds of students that throng every academic building; they look full, and they are. *Québec's* population of university students amounts to 250,000 in round numbers, of whom 85 per cent are enrolled in regular degree programmes. That implies a rate of participation in higher education of roughly 40 per cent of *Québec's* young people — not high for Canada, but creditable by most other standards.

But French-speaking *Québec's* great embarrassment is that at McGill, the most distinguished of its universities, the language of instruction is still English. Indeed, McGill is the chief source of high-level skill, responsible for roughly a third of the 600 or so who graduate with PhDs in *Québec* each year. Moreover, McGill remains a dominant source of physicians and of medical research, despite the emergence of Toronto (the provincial capital of Ontario) as a second string to Canada's bow.

The present array of university institutions in *Québec* is in itself a guide to the history of separatism. McGill University was founded in 1821 almost by accident, when the will of a fur-trader of that name made the founding of a college a condition of a legacy of land and cash; it now has 30,000 students. Next came Bishop's University, an Anglican foundation (in 1843) sited near Sherbrooke, which remains largely a liberal arts college with 2,500 students. Laval University, half an hour's drive west of *Montréal*, is by contrast a Catholic Foundation (chartered as a university in 1853) whose roots lie in the seventeenth century and the seminary founded by the first Bishop of *Montréal* after the French occupation of *Québec*.

The University of *Montréal* is an out-

growth of Laval University, which became independent in 1920 and which has grown to 40,000 students. The *Québec* government's institution-building is represented chiefly by the formation in 1954 of Sherbrooke University and the University of *Québec*, formed in 1968 along the lines of a state university system in the United States; there are six constituent universities, the largest of them (with 40,000 students) is the *Université de Québec à Montréal*.

All *Québec's* universities are in the last resort supported by the provincial government through the Ministry of Higher Education and Science. Students may be awarded bursaries, but otherwise subsist on loans from the government. Apart from capital spending on new buildings, the systems costs the provincial government roughly C\$2 billion a year. The mood among officials is that they will have to raise the funds to finance whatever expansion is driven by the population's inclinations in the years ahead.

The explanation is that the government of *Québec* has made skill a cornerstone of its industrial and social policy. But there is a sense in which this is an uphill task. The output of people with PhD degrees nearly doubled during the 1980s (to 700 in 1990) with a disproportionately fast growth in science and engineering (from 48 per cent to 56 per cent of the total).

Even so, *Montréal* is forever making unfavourable comparisons with Ontario: per head of population, Ontario produces 20 per cent more PhDs a year than *Québec*, but the belief is that the gap is narrowing.

Meanwhile, *Québec* sets great store by its indigenous production of people with master's degrees, usually the result of two years' graduate study. The provincial uni-

versities graduate more than 5,000 people a year at this level, although only a quarter or so of them are qualified in science or engineering.

The university system is importantly stiffened by the way in which research institutes and centres have been scattered through it. These range from the way in which the once independent engineering school, the *Ecole Polytechnique*, has absorbed into the University of *Montréal*, to the Macromolecular Science and Engineering Research Centre at Laval (with an 80-strong staff and more than 40 graduate students) and the Cognitive Research Centre at McGill two-thirds as big).

One chip on *Québec's* shoulder is that the federal government has been less generous towards *Québec* than to Ontario in the siting of federal research establishments. Although the two states are roughly equal in population (at roughly a quarter of the total for Canada), *Québec* has less than an eighth of federal research employees within its boundaries. But the federal Biotechnology Research Centre is conspicuously located in *Québec*.

Even so, the result of all this is that *Québec's* research community is quite substantial. The government reckons that there are 15,000 people working in industrial research and development and perhaps 10,000 (including professorial staff) engaged on research at the universities. In short, the numbers are great enough for the research community to regard itself as a substantial influence within the broader body politic, as well as an instrument for safeguarding its own future.

What should that be? *Québec's* university system has inevitably grown haphazardly, and has consequently a vast in-built diversity. There is good reason to believe that this meets the needs of the young people of the province, but the newer institutions need time in which to become maturely confident in teaching and research. And there is one respect in which the pressures of *Francophonie* are a handicap at this stage; *Québec's* universities are likely, because of the mere existence of the separation problem, to be denied the benefits of cosmopolitanism that healthy university systems need. Should *Québec* seek to avoid this risk by recruiting a steady trickle of academics from other French-speaking countries?

The size of the system also matters. Given people's specialization, it is inevitable that their academic colleagues and competitors will usually be outside *Québec*. Come separation or not, people expect that these links will persist.

The role of McGill in a separate *Québec* is something else again. The entering student body is split roughly 2 to 1 between those with English and French as their mother tongues, but the French-speaking proportion is declining slowly. □



McGill University — dominant but less French than some

The benefits of research

Researchers in *Québec* walk tall. For are they not, together with entrepreneurs, at the centre of the provincial government's affections? That goes without saying. Whether they are adequately supported is another matter, but at least the government of *Québec* has made an important gesture, setting up channels separate from those of the federal government for supporting research.

Canada as a whole, for the size of its population, is well placed in the world's league tables, linked with countries such as Austria and Italy in spending about 1.5 per cent of gross domestic product (GDP) on research and development. Total federal spending in the field has been kept at C\$6 billion in a year of fiscal retrenchment (beginning on 1 April); although the federal government has cut back its investment in the US Space Station and has abandoned altogether a scheme to build a kaon factory at the site of its Tri-University Meson Facility (TRIUMF) in British Columbia, spending on project science has been maintained. For as in *Québec*, so in Canada as a whole, research is now regarded as the road leading to prosperity.

The federal government's arrangements for supporting civil research involve three research councils covering the natural sciences and engineering, medicine and the social sciences and humanities and the Natural Research Council (NRC), now largely concentrated on support for industry, chiefly through its management of federal laboratories. *Québec* reckons to get less than its fair share from the last of these sources.

Comparisons are invidious, but comparisons with Ontario are commonplace in *Québec*. And the plain truth is that total research and development spending in Ontario, at roughly C\$5 billion, was (in 1990) roughly twice that in *Québec*, although the population of Ontario (roughly 10 million) is only half as much again as that of *Québec*. Much of the difference is explained by the greater concentration of NRC and other laboratories in Ontario, which appears to have collared 60 per cent of federal in-house research, but industrial spending on research and development is also twice as great in Ontario as in *Québec*.

Yet *Québec* reckons to be competing successfully for its share of research council funds. *Québec* researchers now routinely win more than 28 per cent of federal research grants, compared with Ontario's 34 per cent; *Québec*'s share of Medical Research Council grants actually exceeds that of Ontario (33 per cent of the total, compared with 29 per cent). But the

provincial government seeks further strengthening of the research effort.

Québec's arrangements for supplementing federal spending have their roots in the 1950s, and now consist of a number of "*Fonds*" for supporting research in fields such as health research (see box below) so as to complement rather than compete with federal research council spending. Although the *Fonds* will occasionally support projects in the familiar sense, it will more commonly support infrastructure development (often instruments or technical assistance), travel and, invaluable, the appointment of people on short-term fellowships.

The consequence of these developments and the general enthusiasm is that *Québec*'s direct support for research, which exceeds C\$200 million a year in total, has grown rapidly (multiplying by more than three in the 1980s). The amounts of money spent do not differ much between *Québec* and Ontario, but the *Québec* government's spending is a larger and thus a more influential proportion (some 9 per cent) of a smaller total.

Building on institutional success

Genest says that, at the outset, he was appalled that French-speaking researchers habitually failed to apply for federally available research funds, taking the view that the inherent virtues of their projects should be recognized and that the authorities should "subsidize them" unsolicited. He believes that the mere existence of *FRSO* has helped to change that view, and to exorcise the sense of neglect that it invited. From its modest beginning, *FRSO* has grown to have an annual budget exceeding C\$60 million a year.

Genest has been a formidable institution-builder in *Québec*. He retired only two years ago as director of the *Institut de recherche clinique de Montréal* which he founded 20 years earlier, and which now houses 30 PhD students in a joint programme with McGill's department of experimental medicine.

Like most other researchers, Genest regrets that separation has become the live issue it now is. He says that it would have taken very little in the way of concession to *Québécois* to have avoided the confrontation on which "extremists" appear bent.

Research set great store by the provincial government's *Fonds*, chiefly on the grounds that they make it possible for research groups to compete successfully for federal funds when they would otherwise be hampered by lack of equipment or departmental spending money. They particularly applaud the grants of up to C\$40,000 that may be awarded (at least by the *Fonds de recharge en santé du Québec* to new faculty recruits to get stuck into research.

By what means can *Québec*'s commitment to research be judged? There are, of course, few objective indicators of the economic value of research; increases of GDP are expected to lag behind research investment by decades, and can be unambiguously linked with them only by students of economic history, working with anecdotal evidence.

Otherwise, it is always possible to count published scientific papers. By that yardstick, three universities predictably dominate *Québec*'s research output McGill (38 per cent), *Montréal* (23 per cent) and Laval (15 per cent). The *Université de Québec à Montréal* is a poor fourth with under 8 per cent.

What rationality would require of *Québec*, at this stage in its constitutional history, is a serious appraisal of the likely long-term economic value of its commitment to research. The potential benefits are not simply those of the patents applied for, or the degree to which *Québec* companies are directly helped to innovation by research spending, but the less tangible benefits of the quantity of skill that *Québec* is generating for itself each year. But this crucial question is not much discussed. It is likely to be answered as if it were an article of faith.

Thus the way ahead for *Québec*, separation or not, is that the amounts of money available through the provincial government's *Fonds* should be even more rapidly increased, to the point at which those in charge of them discover that the quality of the demands upon them is declining. That way, given the difficulty of proving retrospectively that investment in research does lead to prosperity, *Québec* might at least be able to put the proposition to an empirical test.

The best-known, and the oldest, of *Québec*'s provincial research *Fonds* is the *Fonds de recherche, en santé du Québec* (*FRSO*), whose founder, Dr Jacques Genest, used C\$100,000 (mostly from the provincial government) to begin what was at first the *Québec* Medical Research Council. (The name was changed a few years ago, when the government chose to put it on the same footing as the *Fonds pour la formation des chercheurs et l'aide à la recherche*, *FCAR*, and the *Fonds de développement technologique*, *FDT*.)

What are the risks?

What would be the consequences of the constitutional separation of *Québec* for science in *Québec* itself and in Canada more generally? Everything would depend on the manner of the separation, on the fine print of an eventual agreement between *Québec* and the eleven other provinces that constitute Canada now. One of the more curious of present circumstances is that the scientific community seems to have given little thought to what separation would entail in practice.

That is understandable. Whatever their mother tongue, researchers appear to be either against separation or lukewarm about it. Moreover, although the political process could be quick, actual separation must be delayed. A decisive result in a referendum next year could set constitutional talks in train before the end of 1995, but nobody expects that process to be completed quickly.

Meanwhile, there is a sense in which to discuss in detail what separation entails is to tempt Providence, as if talking about separation will make it happen. But the usual working hypothesis is that, in the conduct and funding of research, separation would bring no change. People suppose that the working relationships they have already established with groups elsewhere in Canada, or further afield, would survive. Some even suppose that the federal research councils would survive separation intact.

What would inevitably change is the emphasis given to the French language. The complaint in the 1960s was that *Québec* research languished because applications for funds were not properly understood. Even now that federal Canada is bilingual, traces of this sense of injectice persist. Some in *Quebec* complain that, when applications are now sent to supposedly bilingual referees elsewhere in Canada, they are often imperfectly understood and, as a consequence, are inadequately assessed. In the nature of things, there is no way of distinguishing between that state of affairs and the more general indignation that the doings of referees stimulate.

For a separate *Québec*, there is no reason to suppose that the mechanics of the refereeing process would be a greater handicap than it is now. The international research community, in Canada and elsewhere, would presumably be as ready then as now to read colleagues' manuscripts and grant applications. Moreover, the crucial quality of a referee is that he or she should be able to understand the science. Merely understanding the languages usually a secondary consideration.

A more serious consideration is that a separate *Québec* might be tempted (or even forced by its constitution) to narrow

the circle of those whose opinions help in making decisions. Lively though *Québec's* research community may now be, it is far from self-sufficient. Even Canada's federal research councils include scientists from elsewhere on their advisory committees (some of whom are from France and Belgium). A separate *Québec* would, in its own interests, even more vigorously follow that course.

A much more serious question is whether a separate *Québec* would run its own mechanism for the funding of research. On the face of things, it is improbable that a provincial government that makes research a central pillar of its policy would continue to contribute funds to central research councils so that they could reallocate them among research groups throughout Canada. In other words, *Québec's* present arrangements for supplementing the support offered by the federal research councils seem destined to grow into a separate *Québec's* machinery for research support.

Researchers have good cause to fear that such a development would destroy the plurality they enjoy at present. A more substantial difficulty is that the academic function of the *Fonds* which at present support research in the natural sciences, health research and technology (see previous page) as well as in the social sciences and the humanities, is to enable selected departments or research groups to compete more effectively for federal research funds. The awkward question is whether, without conflict of interest, it would be possible for unified research councils both to single out candidate recipients of research grants and then to award those grants objectively.

These difficulties, about peer-review and funding mechanisms, but especially the second, would apply with almost equal force to Canada without *Québec*. Although some researchers elsewhere would welcome the illusion that separation would rid them of the present need to be bilingual, the geographical gulf between Ontario and British Columbia would then seem even greater than now.

Some in *Québec* raise a further difficulty, saying that the provincial government's rhetoric about its commitment to research has not been fully reflected in the annual budgets of *Québec's* independent research *Fonds*. All governments are faced with the difficulty of knowing how much research support is enough, but statistics support the complaint that the *Fonds* could usefully spend more.

Although *Québec's* direct support of academic research grew fourfold in 1980s (to about C\$120 million a year), both the value and the proportion of federal research grants awarded to researchers at

Québec universities and institutions are still increasing. In other words, it seems that the *Fonds* have not yet exhausted the supply of university departments and groups able to compete successfully with others elsewhere in Canada. Even if separation never happens, the provincial government might usefully be challenged on that score.

But the most serious danger of separation would be that with which *Québec's* university system would be faced. Much of it is relatively new, with the consequence that there has not yet been time for some institutions to strike a balance between teaching and research, McGill University, with 30,000 students, has 1,500 faculty members and was responsible in 1991 for 38 per cent of *Québec's* output of scientific publications. The *Université de Québec à Montréal*, by contrast, has 40,000 students, 1,000 faculty members and produced 7.6 per cent of *Québec's* research publications in the same year.

The purpose of this comparison is nothing other than to suggest that, for reasons to do with the balance between teaching and research, not all components of *Québec's* university system are equally active in research. But that, in turn, implies that the effective mass of *Québec's* academic research enterprise is less than simple head-count would suggest. Yet even by that yardstick, it is small — too small in many fields to be a critical mass. Researchers seem ready to acknowledge that; it is not clear whether politicians bent on separation agree.

These are the moderate alarms occasioned by the prospect of separation. There are, of course, other less moderate alarms, not least that a separate *Québec* would be a chauvinistic *Québec*, culturally and politically more distant from the remainder of Canada than from, say, the United States or even France. No doubt the remainder of Canada would abreact to such a development, with disastrous consequences for general civility.

That is one reason why the general detachment of the research community from the issue of whether separation becomes a reality is difficult to understand. For *Québec*, the project has the attractiveness of an adventure without obvious or immediate risk. The foreseeable dangers are more distant, while the chance of blundering into crisis is discounted.

For the rest of Canada, the outlook is very much the same: the hazards of separation would depend on how separate a separate *Québec* would be. But Canada as a whole seems not yet to have appreciated how the clamour for separation, like the recent wave of immigration from South East Asia, has drawn attention to the cultural diversity of a vast country better known for being empty of people, even of interest. Is that not worth struggling to keep?

John Maddox

Sun's encounter with star predicted

The Solar System proper seems safe from external calamity for the next 50 years, but there is a danger that a close encounter with α Centaurus A/B could shake loose a shower of comets on the Earth.

THERE is, no doubt, widespread relief that Comet Swift-Tuttle is not now believed likely to collide with the Earth on its next-but-one return, in 2126 (see *Nature* 367, 681; 1994). But before rejoicing gets out of hand, the world should know of a potentially more serious threat to its peace of mind — the prospect that there will be a veritable shower of comets crossing the Earth's orbit in the million or so years following the next close encounter of the Solar System with another star, likely to take place 30,000 years or so from now.

This calculation is due to Robert A. J. Matthews, who divides his time between research and as a science writer for the London newspaper *The Sunday Telegraph*, and is published in the current issue of the *Quarterly Journal of the Royal Astronomical Society* (35, 1–9; 1994). Matthews has worked out the probable motion of some of the stars known to lie within 5 parsecs of the Sun. His immediate objective is to tell which of several is likely to succeed Proxima as the Sun's nearest neighbour. The arithmetic is not as simple as might be expected.

Proxima itself is a dwarf star, perhaps 10 per cent the mass of the Sun. From time to time, there are flares on its surface, suggesting (but not proving) that it is relatively young. Its present distance from the Sun, inferred from parallax measurements (of its position relative to more distant stars from opposite ends of the Earth's orbit) is 1.295 parsecs. (One parsec is the distance at which the Earth's orbit would subtend an angle of one second of arc; numerically, it is 206,000 astronomical units, where one AU is the radius of the Earth's orbit about the Sun.)

Listing nearby stars is not child's play. Brightest is not necessarily nearest; only parallax can settle that. But there seem to be 58 stars within a 5 parsec radius of the Sun. To know the present velocity of a star, two other measurements are required; its radial velocity and its angular motion across the sky, otherwise its proper motion.

Sadly, these measurements of Proxima cannot be used to extrapolate its future motion relative to the Sun because of a pre-existing difficulty, that of knowing whether Proxima is a free agent or, otherwise, the third member of the binary system known as α Centarus A/B, a binary star in which each component is as massive as the Sun and only a little further away than Proxima at 1.335 parsecs.

Published data, according to Matthews, cannot decide whether Proxima is free or bound. Whatever the truth, it can only be marginally captive. The distance between Proxima and α Centaurus A/B leads to a lower limit for the semi-major axis of its presumed orbit about the binary of 13,000 AU; similarly, the period of rotation must be at least one million years.

For what it is worth, Matthews says he has been convinced by unpublished data from the European Southern Observatory's measurements of radial velocities at La Silla in Chile that Proxima is indeed bound, which raises an interesting conundrum. The two stars of the binary are main-sequence stars, probably more ancient than the Sun. One possibility is that they began life as a triple star, the other is that the binary has more recently captured Proxima. Given the flares on Proxima, capture is the more likely origin.

Either way, the notion that Proxima is indeed bound into a triple star makes it possible to correct its radial velocity and to make a more accurate extrapolation of its motion than would otherwise be possible. The upshot is that Proxima will be closest to the Sun 26,700 years from now, at a distance of 0.941 parsec. What effect, if any, will that close passage have on the Solar System?

This is where the binary comes into its own. Wherever Proxima is in the future, α Centaurus A/B will not be far away. Moreover, the mass of the binary is some 20 times that of Proxima and thus likely to be much more influential. But for the record, it is worth noting Matthews' conclusion that during the next 50,000 years at least three other stars (Barnard's star, Ross 248 and AC + 79° 3888) will, in that order, come within a parsec or so of the Sun. Indeed, Matthews says "we are approaching a time peculiarly rich in stellar encounters" when near-approaches to the Sun will be ten times more frequent than expected from the distribution of stars in the Galaxy.

The effects of α Centaurus A/B on the Solar System will not at first be dramatic. For one thing, there will be no perceptible disturbance of the planets, which all lie comfortably within 50 AU of the Sun. Such effects as there will be will involve the so-called Oort cloud of proto-comets believed to lie beyond the planets but within a radius of less than half a parsec (say, 100,000 AU) of the Sun. The passage of α Centaurus A/B could set calamitous events in train there.

That the Oort cloud, which commands

varying degrees of belief and scepticism, exists is no longer in dispute; the difficulties lie in the estimates of the number of proto-comets it contains and, more pertinently for the effects of the near-encounter with α Centaurus A/B 30,000 years or so from now, the distribution of proto-comets with distance from the Sun. There is no shame in that, as can be told from how the population of the Oort cloud was first painstakingly estimated from records of comets in nearly, but not quite, unbound (or 'hyperbolic') orbits.

The physics of the near-encounter with α Centaurus A/B is simple. The star would provide each proto-comet with a mechanical impulse, which can be calculated from the trajectory of the star and the supposed distance of the proto-comet. Matthews concludes that the projected approach of α Centaurus A/B will shake loose 200,000 comets from the Oort cloud, providing them with at least the potential to move within the Earth's orbit around the Sun. But a substantial fraction of those would be ejected from the Solar System altogether by unfavourable encounters with Saturn and Jupiter, although others would become more or less permanent Earth-crossing comets. It is worth noting that the total Oort population, estimated at a small multiple of 10^{12} proto-comets, would not be substantially diminished by the encounter.

Luckily, the timescale of this process is not short. The orbital period of an Earth-crossing comet created by impulsive perturbation of a proto-comet in the Oort cloud at 100,000 AU would be about a million years. Other objects able to threaten the Earth by impact, perhaps by interaction with Neptune and Uranus or Saturn and Jupiter would become perils only after an interval ten times as great. In other words, there is time in hand.

That raises an absorbing question. The close encounter with α Centaurus A/B 30,000 years from now will take place only after the passage of 100 times as much time as there has been since Galileo first saw the motion of the moons of Jupiter. It is not easily conceivable that the people then practising astronomy will not have figured out in detail what the Oort cloud consists of. It is not inconceivable that its proto-comets will have been recorded in some database. In other words, the prospect of calamity will be an operational problem. Is that what Spring holidays are about?

John Maddox

Bioremediation comes of age

Richard P. J. Swannell and Ian M. Head

TECHNOLOGY proven in the laboratory can take a surprisingly long time to be used in earnest. This is certainly the case for oil spill bioremediation, as the idea that the biodegradation of oil in the environment might be stimulated by the addition of inorganic nitrogen and phosphorus was first mooted 21 years ago¹. It is fitting that one of the original proponents of this idea, Ronald Atlas, should be part of the team that has now demonstrated convincingly that simply adding inorganic

oil by converting hydrocarbons largely to recyclable products (including carbon dioxide, water and cellular biomass) could greatly reduce the long-term effect of a spill on shoreline biota. A flurry of research generated new products which were studied on a range of coastlines^{3,4}. But although bioremediation did increase oil biodegradation rates on contaminated beaches it was not an instant cure-all³⁻⁶. As positive reports of bioremediation have been balanced invariably by failures,

remediation has in fact been effective in the Alaskan clean-up operations, Bragg *et al.* have come up with an alternative way of assessing the degree of oil biodegradation on polluted beaches⁸. This method was initially developed by petroleum geochemists to characterize oils found in subterranean reservoirs. It uses the concentration of 17 α (H),21 β (H)-hopane as the non-biodegradable oil component against which to index the biodegradation of the more labile hydrocarbons present in the oil. As Bragg *et al.* show, this approach provides the statistical verification that bioremediation has been successful.

Perhaps a more significant finding is the positive correlation between concentration of nitrogenous nutrients in sediment pore waters and the extent of biodegradation. Biodegradation was shown to be effective only on beaches where high levels of pore water nitrogen were maintained. This implies that for successful oil bioremediation, monitoring of nutrient concentrations in the field is necessary.

The data obtained by Bragg and his colleagues allow them to model the bioremediation process, something that should aid the development of response strategies and reduce the need for field trials. Knowing the sediment oil loading and the proportion of poorly degradable polar compounds present in the oil (the higher this proportion the less likely bioremediation is to succeed), environmental scientists should be able to use these models to assess how effective a bioremediation strategy is likely to be.

Although bioremediation has proved its worth on the porous cobble shorelines of Alaska, the book is far from closed. Will the strategy be equally effective on mudflats, fine sands or salt marshes where oxygen may be in short supply⁶? Oxygen availability, one key feature that is absent from the models of Bragg *et al.*, has proved to be an important factor not only for the fate of oil spilled on beaches⁹ but also in instances such as the *Braer* spill last year (off the coast of Shetland, UK),

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Clean-up operation on the coast of Knight Island, Alaska, after the *Exxon Valdez* oil spill.

nutrients can stimulate the natural rate of oil biodegradation on polluted shorelines. With the compelling evidence now described on page 413 of this issue², this technique for the treatment of oil spills has finally come of age.

The history of bioremediation has been beset by many false dawns. In the mid-1970s, its use to treat marine oil spills in particular was evaluated both in the laboratory and in the field. Some trials were successful, but the enhancement of degradation was never enough to provide a practical response to a floating oil slick.

In 1978, the *Amoco Cadiz* spill of 220,000 tonnes of oil off the coast of Brittany rekindled interest in bioremediation, particularly for shoreline clean-up. Biodegradation was identified as the main natural process by which non-volatile hydrocarbons are removed from the environment. Hydrocarbon-degrading microorganisms are ubiquitous in most ecosystems³⁻⁷, and over 30 different genera of oil-degrading bacteria and fungi have been identified⁶. So the hope was that enhanced microbial decomposition of

many specialists have remained unconvinced of its attributes as a counter-measure against oil spills⁵.

Attitudes changed as a result of the clean-up operations following the 33,000-tonne spill of crude oil from the *Exxon Valdez* in Prince William Sound, Alaska, in 1989. Part of the response strategy for this oil spill was the use of an oleophilic fertilizer (Inipol EAP22) to stimulate biodegradation of the crude oil. Within two weeks, it was obvious from looking at the polluted beaches that much of the oil had been removed, but it was difficult to verify this simply by measuring hydrocarbon concentrations⁷. The established method for monitoring the degree of biodegradation of crude oil, by measuring the ratio of readily biodegradable C₁₇ and C₁₈ straight-chain alkanes to more recalcitrant branched-chain analogues (pristane and phytane), proved unhelpful because the indigenous microorganisms in Prince William Sound biodegraded pristane and phytane almost as fast as their corresponding straight-chain alkanes⁷.

To provide clear evidence that bio-

Vanessa Vick/SPL

1. Atlas, R. M. & Bartha, R. *Envir. Sci. Technol.* **7**, 538-541 (1973).
2. Bragg, J. R., Prince, R. C., Harner, E. J. & Atlas, R. M. *Nature* **368**, 413-418 (1994).
3. Sveum, P. & Ladousse, A. in *Proc. 1989 Oil Spill Conf.* 439-446 (American Petroleum Institute, Washington DC, 1989).
4. Lee, K. & Levy, E. M. in *Proc. 1989 Oil Spill Conf.* 479-486 (American Petroleum Institute, Washington DC, 1989).
5. Owen, D. in *Proc. 14th Arctic and Marine Oil Spill Prog. Technical Seminar* 119-130 (Environment Canada, 1991).
6. Prince, R. C. *Crit. Rev. Microbiol.* **19**, 217-242 (1993).
7. Pritchard, P. H. *et al.* *Biodegradation* **3**, 315-335 (1992).
8. Douglas, G. S. *et al.* in *Contaminated Soils: Diesel Fuel Contamination* (eds Icosteck, P. T. & Calabrese, E. J.) 1-21 (Lewis, Boca Raton, 1992).
9. Lee, K. & Levy, E. M. in *Proc. 1991 Oil Spill Conf.* 541-548 (American Petroleum Institute, Washington DC, 1991).
10. Wang, X. *et al.* *Envir. Sci. Technol.* **24**, 1086-1089 (1990).
11. Jones, D. M. *et al.* *Int. J. env. analyt. Chem.* **24**, 227-247 (1986).
12. Rosenberg, E. *et al.* *Biodegradation* **3**, 337-350 (1992).

where much of the oil sank to the sea floor.

Furthermore, aromatic compounds present in crude oil and diesel fuel may be readily biodegradable initially, but there is evidence that over time a proportion of these compounds becomes tightly adsorbed to the organic matter in soils and sediments, and thus becomes less available to hydrocarbon-degrading microorganisms^{10,11}.

There is certainly scope for the development of bioremediation products that will deliver nutrients more efficiently¹². But the debate continues over whether competent microorganisms should be deliber-

ately introduced, particularly where there is a paucity of indigenous hydrocarbon degraders at a spill site. It may be operational factors such as these, rather than definitive scientific proof of its efficacy, that will determine the widespread adoption of bioremediation as a treatment for spilled oil. □

Richard P. J. Swannell is at Biotechnology Services, AEA Technology, 353 Harwell, OX10 0RA, UK. Ian M. Head is at the Department of Fossil Fuels and Environmental Geochemistry at the University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.

T-CELL SELECTION

A little of what you fancy . . .

Philippa Marrack and David C. Parker

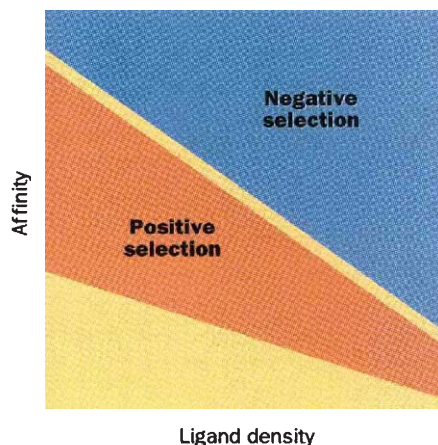
THOSE who complain that the immune system is too complicated to understand should take heart. One of the great riddles of immunology — how T cells are selected in the thymus to recognize foreign antigens in the context of self alleles of the major histocompatibility complex (MHC) — is giving way to experimental investigation, as reported in two papers published earlier this year in *Cell*^{1,2}.

Those antigen receptors of T cells (TCRs) that are made up of α - and β -chains recognize foreign proteins only in the form of short peptides bound to grooves on cell-surface proteins, the products of the MHC. The genes encoding MHC molecules are remarkably polymorphic, and in general, a T cell will not respond to its antigenic peptide bound by the wrong allele of MHC, a phenomenon called MHC-restriction. Like the other antigen-specific receptors of the immune system, the genes for $\alpha\beta$ TCRs are created by genetic rearrangement and imprecise joining. This process produces a huge number of different TCRs, but only a small proportion of them appear on the mature T cells in any given animal. Precursor T cells bearing receptors which can bind well to self peptides on self MHC alleles die in the thymus, a process called negative selection which leads to self tolerance.

Mysteriously, however, precursor T cells cannot mature unless their receptors bind and recognize self alleles of MHC proteins on cells in the thymus. This phenomenon is known as positive selection, and its contradictions with the processes of tolerance have long troubled immunologists. Simply put, how can T cells be both positively and negatively selected on self MHC alleles in the thymus? Although many explanations of this paradox have been proposed, recent evidence favours the 'avidity hypothesis'

— the idea that developing T cells are positively selected if their receptors are barely, almost imperceptibly, engaged with peptide plus MHC ligands in the thymus, and will die if their receptors are strongly engaged.

The new studies come from the laboratories of Michael Bevan¹ and Susumu



Receptor occupancy, a function of the intrinsic affinity of the TCR for its peptide/MHC ligand and the ligand density on the selecting cells, may determine the fate of developing T cells in the thymus. (Figure after an unpublished figure by M. Bevan.)

Tonegawa², and they make use of two types of gene knockout mice deficient in surface expression of class I MHC proteins. In each type of mouse, MHC proteins can be loaded with peptides supplied by the investigator, replacing the many different self peptides provided by the thymus cells themselves in normal mice. In organ culture of fetal thymuses from these knockout mice, maturation of class I-restricted T cells depends on addition of exogenous peptides. The authors of the current papers had previously shown that single peptide/MHC combina-

tions can positively select thymocytes, but not as efficiently as multiple peptide/MHC ligands. From this it was inferred that positive selection is peptide-specific as well as MHC-specific, but the question of which peptides select which TCRs was not addressed.

To test the avidity hypothesis for positive selection, the two laboratories^{1,2} used knockout mice transgenic for rearranged TCR genes. In these mice, nearly all the thymocytes express the transgenic TCR, so that one can easily follow the fate of cells expressing that TCR. In one case (Bevan's group), the TCR was specific for a peptide from chicken ovalbumin bound to the class I MHC protein, K^b. In the other (Tonegawa's group), the TCR was specific for a peptide from the mouse lymphocytic choriomeningitis virus (LCMV) plus the class I protein, D^b.

The papers show that T cells bearing the transgenic TCRs do not mature in organ cultures of fetal thymuses from the class I defective animals unless particular peptides are added. In both instances the successfully selecting peptides are related to the antigenic peptide for which the TCR is specific, but they would be expected to react less effectively because of structural differences from the antigenic peptide (low affinity) or low representation bound to the MHC molecules of the thymic cortical epithelium (low ligand density).

For example, T cells bearing the receptor specific for the ovalbumin peptide are selected by close relatives of the ovalbumin peptide but not by unrelated peptides, which were chosen to bind just as well to K^b. In this case the successful peptides are known to be receptor antagonists, low-affinity analogues of the antigenic peptide which competitively block T-cell activation by the antigenic peptide (see box overleaf). Both groups found that higher densities of a positively selecting peptide/MHC ligand would cause target T cells to die rather than mature. For example, a positively selecting ovalbumin antagonistic peptide killed T cells bearing the transgenic TCR in wild-type thymuses that express normal levels of peptide/MHC ligand. Likewise, high concentrations of the LCMV peptide caused the death of the transgenic T cells whereas low concentrations of peptide enabled positive selection.

These experiments show that the paradox of positive selection versus tolerance may be a simple matter of receptor occupancy, a function both of the affinity of the TCR for its peptide/MHC ligand and of the concentration of the ligand (see figure). But they have not solved all the problems, and the papers themselves contain some contradictions.

For example, in the LCMV studies, immature thymocytes were selected by low concentrations of a peptide which is

Antagonists and partial agonists

A FAVOURITE technique for studying T-cell specificity has been to offer a set of T-cell clones specific for a particular antigenic peptide a set of synthetic peptides that differ in one or more positions. Such experiments have shown that amino acids at some positions in the peptide are important for binding to the MHC molecule while others affect only recognition by the T cell, and are thought to contact the T-cell antigen receptor. Conservative substitutions in most of the residues not essential for MHC binding are still recognized by most T-cell clones, but higher concentrations of the cross-reactive peptide analogue are frequently required to elicit an equivalent response; presumably, the analogues have a lower affinity for the T-cell antigen receptor, and so a higher concentration of peptide/MHC molecules on the antigen-presenting cell is required to give an equivalent level of antigen receptor occupancy.

This picture has been complicated by the finding that some peptide analogues have antigen-specific effects on T cells that are different in kind from those of the antigenic peptide. For example, some analogues act as specific competitive antagonists for a T-cell clone: they block all measurable T-cell responses to an

antigenic peptide when present in excess over the antigenic peptide⁵. And other analogues cannot induce interleukin-2 production and T-cell proliferation at any concentration, but can still elicit other responses from a T-cell clone, such as cellular cytotoxicity or T-cell help for B-cell responses⁶. Some of these peptides with partial agonist properties are also antagonists when presented together with an antigenic peptide⁷.

The discovery of partial agonist and antagonist peptides suggests that the T cell has some way to measure the *quality* of the interaction of an individual antigen receptor with an MHC/peptide complex, as well as overall receptor occupancy. Perhaps the half-life and precise alignment of the individual TCR/peptide/MHC complex determine the kind of signal by limiting the assembly of a signalling complex (for example, MHC dimerization or CD4 association).

Antagonism could provide a safe solution to the paradox of positive and negative selection in the thymus. The hypothesis is that a self peptide that positively selects a T cell in the thymus would interact with the T-cell receptor in the same range of affinity or in the same way as an antagonist peptide for a mature T

cell. In this way, positive and negative selection could occur on largely non-overlapping sets of self peptides. More importantly, autoimmunity would be neatly avoided were a mature T cell to encounter its selecting self peptide on some vulnerable tissue in the body at higher concentrations than in the thymus, because the self peptide would antagonize rather than stimulate the T cell.

Superficially, the results of Bevan and colleagues¹ support this view, because they found that the peptides that positively select a particular TCR are a subset of the peptides previously found to be antagonists for that TCR in a mature T cell. But the correlation may be more apparent than real, and reflect the levels of representation of the tested peptides on the thymus epithelium. If higher levels of representation could be achieved, it is possible that more distantly related peptides might positively select, as reported by Tonegawa and co-workers². Moreover, both groups found that the same peptide can act in both positive and negative selection for a particular T-cell receptor, depending on peptide/MHC concentration on the surface of the selecting cell.

P.M. & D.C.P.

strongly antigenic for the TCR in a mature T cell line. Pamela Ohashi's group, using the same LCMV TCR but the type of knockout mouse used by Bevan and colleagues, has also found that the antigenic peptide can positively select its TCR (ref. 3). The native ovalbumin peptide, on the other hand, did not select its TCR, even when it was added at very low concentrations to the thymus organ cultures. So there may be an intrinsic affinity threshold above which only negative selection is possible. Also, in the ovalbumin peptide studies, some antagonist peptides could positively select whereas others, which appeared to be equally good antagonists, could not. It is possible that mature T cells and thymocytes can discriminate between ligands other than by simple receptor occupancy (see box). Another study provides evidence that an antagonist peptide may antagonize positive selection in fetal thymic organ culture⁴, a possibility inconsistent with any simple receptor occupancy model. Perhaps a better understanding of the mechanisms of antagonism will yet shed light on positive selection.

Although the new data show clearly that positive selection is peptide specific and dependent on TCR affinity and ligand concentration, the degree of specificity or degeneracy in positive selection in a normal thymus remains to be determined. For example, Tonegawa's group has re-

ported that a combination of two particular exogenously added peptides selected as many heterogeneous T cells (in non-TCR transgenic mice) as a mixture of the many class I-binding peptides isolated from spleen cells, a result which is difficult to reconcile with the peptide specificity of positive selection demonstrated in the current papers. Thus, the matter of which and how many peptide/MHC combinations contribute to positive selection of a particular TCR in a normal thymus remains open.

There are other outstanding questions, the most important of which concerns the ability of developing T cells to detect very low affinity, low ligand densities with their TCRs. Immature thymocytes bear 5–10 times fewer TCRs per cell than mature T cells, yet they can detect ligands and ligand concentrations which are apparently not seen by mature T cells. Given that the affinities of TCRs for their antigenic peptide/MHC ligands are very low, some 10^{-5} M according to the few published measurements, how do thymocytes discriminate between ligands with affinities much less than this? Is the exquisite sensitivity of immature thymocytes aided by the fact that their TCR/CD3 complex is already partially phosphorylated or that their TCRs are incompletely engaged by CD3? Do immature thymocytes possess as yet undiscovered ligands that could stabilize very weak interactions

between TCR and peptide/MHC? What are the signals for positive selection that require thymic organ culture and cannot yet be reproduced in cell suspensions? Finally, are the rules for positive selection of thymocytes destined to be restricted by class II MHC proteins similar to those for class I?

Nonetheless, our guess is that the problem of positive selection is cracking before our eyes. The phenomenon involves an interaction between the receptors of developing thymocytes and peptide/MHC complexes at affinities or ligand densities too low to drive T-cell activation or thymocyte death. This is truly a case of a little of what you fancy does you good. □

Philippa Marrack is in the Howard Hughes Medical Institute and the Department of Medicine at the National Jewish Center for Immunology and Respiratory Diseases, Denver, Colorado 80303, USA. David C. Parker is in the Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA.

1. Hogquist, K. A. *et al.* *Cell* **76**, 17–27 (1994).
2. Ashton-Rickardt, P. G. *et al.* *Cell* **76**, 651–663 (1994).
3. Sebzda, E. *et al.* *Science* **263**, 1615–1618 (1994).
4. Spain, L. M., Jorgenson, J. L., Davis, M. M. & Berg, L. J. *J. Immunol.* **152**, 1709–1717 (1994).
5. De Magistris, M. T. *et al.* *Cell* **68**, 625–634 (1992).
6. Evavold, B. D., Sloan-Lancaster, J. & Allen, P. M. *Immun. Today* **14**, 602–609 (1993).
7. Evavold, B. D., Sloan-Lancaster, J. & Allen, P. M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Variable but singular

Leslie C. Aiello

IN human palaeontology there are often more opinions and interpretations than there are fossils. Happily, however, the newly discovered *Australopithecus afarensis* skull and other remains from the Hadar Formation in Ethiopia, described by Kimbel, Johanson and Rak on page 449 of this issue¹, will go a long way to settle some of the most heated controversies surrounding the earliest species in the human lineage.

Australopithecus afarensis was established² in 1978 for the fossils from Laetoli, Tanzania, and from Hadar in Ethiopia. The *A. afarensis* hypodigm — that is, all specimens designated under that species name — includes the famous partial skeleton (A.L. 288-1) nicknamed Lucy, and we now know that it spans a period of almost a million years from about 3 million years ago to 3.9 million years ago. The newly discovered skull (A.L. 444-2) from the Kada Hadar Member at Hadar is the youngest known *A. afarensis* fossil, the older date belonging to the Belohdelie frontal (BEL-VP-1/1) from the Middle Awash; the previously tentative attribution of this specimen to *A. afarensis* has been confirmed by its similarity to the new find. We also know that *A. afarensis* occupied a variety of habitats, from the relatively open and dry to wetter, forested conditions³. These interpretations are not contentious.

What has been contentious is the issue of whether all of the fossils included in the *A. afarensis* hypodigm actually represent one sexually dimorphic species. Entangled in this debate is the question of locomotion. Were there major differences in bipedal capabilities between the smaller, apparently more arboreal, Lucy-sized hominids and the larger, apparently more bipedal, individuals as some claim⁴? Does this indicate quite a remarkable sexual difference in locomotor capabilities in our early ancestors or is it evidence for more than one species⁵? Furthermore, were any of the *A. afarensis* hominids good enough on two feet to have made the Laetoli footprints⁶?

At first glance, the new fossils reported by Kimbel *et al.*, and particularly the skull, the humerus (A.L. 137-50) and the ulna (A.L. 438-1a), suggest individuals so large and robust that it would seem difficult to accommodate them in the same taxon as the diminutive Lucy. The skull is the largest yet known for any australopithecine, with a biasterionic breadth of 106 mm, 11.5% larger than the same measurement for A.L. 333-45 (ref. 7). The inferred length of the robust humerus is just under 24% longer than Lucy's humerus and the new ulna is about 22% longer, showing considerably more dimorphism than is the norm in humans where both the humerus and the ulna have a length dimorphism of about 11%.

This degree of dimorphism might not be unlikely for *A. afarensis*, however. Arguments over the detailed morphology of the small and large individuals aside, estimates of body weight show that males of all australopithecine species are very similar in size, between 40 kg for *A. africanus* and *A. robustus* and 49 kg for *A. boisei*, and that the sexual dimorphism in body weight ranged between 66% in *A. afarensis* and 79% in *A. robustus*⁸. *Australopithecus boisei* approaches *A. afarensis* most closely in body weight dimorphism (70%) and also shows a similarly high dimorphism in biasterionic breadth (17.5% between KNM-ER 406 and 407)⁷. It would seem that all australopithecines were highly dimorphic and that the range of size dimorphism in *A. afarensis* is not particularly unusual.

One of the most important of the new finds is the virtually complete ulna. Its degree of shaft curvature and length relative to humerus length are most reminiscent of chimpanzee ulnae, which have many features that are consistent with adaptation for climbing⁹. But the *A. afarensis* ulna lacks the chimpanzee proximo-anterior orientation of the trochlear notch and the very small olecranon, features that are compatible with

terrestrial, quadrupedal locomotion where the body weight is supported on the forelimbs. The mosaic morphology of the *A. afarensis* ulna, together with the heavily muscled and robust humerus, would be ideally suited to a creature which climbed in the trees but also walked on two legs when on the ground. Furthermore, the apparent anatomical similarity between the Hadar ulna and two later robust australopithecine ulnae, one from the Omo (Omo 40-19) and the other from Olduvai Gorge (OH 36), suggests that this forearm morphology persisted in other australopithecine species for well over a million years.

But it is the skull — the first relatively complete example known for this species — that will attract most attention. Not only does it confirm the fundamentally primitive nature of *A. afarensis* cranial anatomy in relation to later australopithecines, but it also provides a wealth of information for functional and phylogenetic inference. One of the most notable

Passing splendour

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Popperfoto

Amorphophallus titanum, the world's largest flower, in bloom at a botanical garden in West Java on 21 March. Size is not the plant's only feature, for it exudes an insect-attracting stench variously described as like that of "a mixture of rotten fish and burnt sugar" or of "a ripe dead rat". Flowering is rare and transitory, occurring every four years and lasting but a few days. So now the show is over. T.L.

1. Kimbel, W. H., Johanson, D. C. & Rak, Y. *Nature* **368**, 449–451 (1994).
2. Johanson, D. C., White, T. D. & Coppens, Y. *Kirtlandia* **28**, 1–14 (1978).
3. White, T. D. *et al.* *Nature* **366**, 261–265 (1993).
4. Stern, J. T. & Susman, R. L. *Am. J. phys. Anthropol.* **60**, 279–317 (1983).
5. Senut, B. & Tardieu, C. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 193–201 (Liss, New York, 1985).
6. White, T. D. & Suwa, G. *Am. J. phys. Anthropol.* **72**, 485–514 (1987).
7. Wood, B. *Koobi Fora Research Project Vol. 4: Hominid Cranial Remains* (Oxford Univ. Press, 1991).
8. McHenry, H. *Am. J. phys. Anthropol.* **87**, 407–431 (1992).
9. Aiello, L. C. & Dean, M. C. *An Introduction to Human Evolutionary Anatomy* (Academic, London, 1990).
10. de Bonis, L., Bouvraïn, G., Geraads, D. & Koufos, G. *Nature* **345**, 712–714 (1990).

avenues it opens is the possibility of assessing the relationship between *A. afarensis* and some of the earlier Miocene apes, such as *Ouranopithecus macedoniensis*¹⁰, that have been suggested to be close to the common ancestry of the chimpanzee and human lines.

The new finds at Hadar, together with other recent evidence from the Middle Awash, Ethiopia³, provide persuasive support for the idea that *A. afarensis* is a single, highly dimorphic species. Any differences in bipedal ability that might be inferred from fragmentary lower limb fossils have also been overshadowed by the recognition that the forelimb of the large, and purportedly more bipedal,

males is most probably as adapted to arboreality as is the forelimb of Lucy⁴. Furthermore, if these hominines were not good enough bipeds to make the Laetoli footprints, there are no other candidates currently known in the record that could have been responsible. It would now take compelling fossil evidence to start the pendulum of opinion swinging back towards the idea that there were several species within the *A. afarensis* hypodigm. Stranger things, though, have been known to happen in human palaeontology. □

Leslie C. Aiello is in the Department of Anthropology, University College London, Gower Street, London WC1E 6BT, UK.

LASER PHYSICS

Scattering for super-radiation

A. Z. Genack and J. M. Drake

THE spatial, temporal and spectral coherence of lasers has made them powerful tools for a dizzying assortment of precise measurements and for the efficient delivery of high optical powers. The lasing threshold is reached when the round-trip gain of the portion of the beam which is

fed back into the amplifying medium exceeds the loss at the laser output couplers. Care is generally taken to ensure that the lasing medium is homogeneous lest the pump or emitted beams be scattered out of the gain medium and their spatial coherence destroyed. As lasing and disorder appear to be incompatible, it would seem to be folly to attempt to promote lasing by deliberately introducing scatterers into a medium. Yet that is what Lawandy, Balachandran, Gomes and Sauvain have done — they report on page 436 of this issue that lasing action is observed from a dye solution when titania particles are introduced, but not otherwise. The solution to the conundrum seems to lie in appreciating the changes in the distribution of paths and in their interactions in the amplifying medium under various excitation conditions.

Some 25 years ago, V. S. Letokhov (*Sov. Phys. JETP* **26**, 835–840; 1968) considered the generation of light in a random medium with gain in the case that the photon transport mean free path, l , is much smaller than the dimensions of the sample, L . He proposed that placing scattering particles in an amplifying gaseous medium could enhance the frequency stability of the emission. As there is no preferred direction for the field, the probability of stimulated emission would not depend on the direction of the atom's velocity, and the frequency of the

super-radiant emission from the system would coincide with the centre of the atomic absorption line. An antithetical limit has been considered recently (P. Sebbah, C. Vanneste and D. Sornette, personal communication) of stimulated emission in a localizing medium in which discrete stationary modes of the scattering medium act as distributed resonant cavities which set the frequencies of the super-radiant emission.

The propagation of light in a scattering medium can be described as a random walk of photons with the direction of each succeeding step determined by the laws of probability. Not surprisingly, the spatial and temporal evolution of the optical intensity is described by the particle diffusion equation. For the case of uniform photon generation within a scattering region, Letokhov solved the diffusion equation with the gain included by simply reversing the sign of the usual loss term. The presence of scatterers increases the path length in the gain medium. Because stimulated emission occurs in all directions, the possibility of the wave scattering out of some narrow gain cone is eliminated. These factors conspire to allow the path length in the gain medium to exceed the super-radiant gain length l_{sr} , in which the probability of stimulated emission is unity.

In practical experiments, the gain in a random medium is not uniform. In the study by Lawandy *et al.*, gain is produced by a pumping laser focused on the face of a cuvette containing the amplifying titania colloid (Fig. 1). A typical path for an emitted photon within the excitation region is shown schematically in Fig. 2. Although the pure dye solution used does not show laser action in their experimental configuration, super-radiant emission does emerge from the front face of the sample when scatterers are introduced at a concentration of about 10^{10} cm^{-3} .

The onset of super-radiant emission occurs when the pump intensity is equal to the saturation intensity of the dye solution, coinciding with a dramatic narrowing of the spontaneous emission bandwidth and a change in the slope of a plot of input against output intensity. The observation of super-radiance signals that typical paths of spontaneously emitted photons are lengthened by the presence of scatterers so that they exceed the super-radiant gain length. It is puzzling, however, that in recent measurements using higher powers of exciting radiation and lower dye concentrations than this, M. Siddiq, C. Liu and R. R. Alfano (personal communication) have observed super-radiant emission in the pure dye solution but find that the introduction of a low density of scattering particles raises the threshold pump power required to produce super-radiance.

What explains these diverse phe-

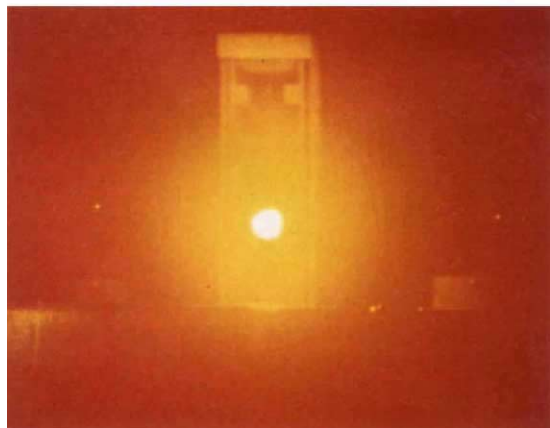
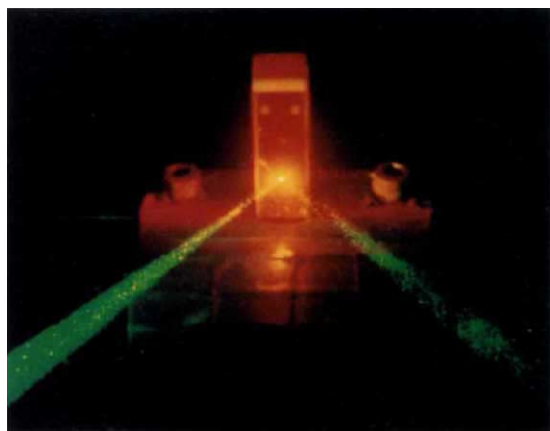


FIG. 1 How a lick of paint can brighten a whole room. Top, the pure dye solution used by Lawandy *et al.*, pumped by green laser light; bottom, with added colloidal titania particles (white paint, essentially). The excitation beams practically vanish in the glow.

N. M. Lawandy/Brown University

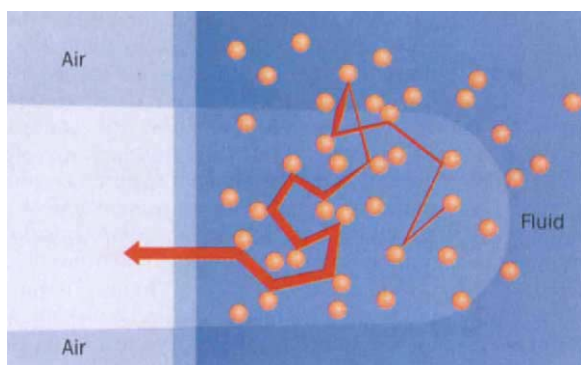


FIG. 2 Possible path for a photon emitted and amplified within a dye medium containing random scatterers. The lighter region indicates the volume strongly pumped by the laser.

nomena, and in particular, what are the conditions for producing compact super-radiant devices in strongly scattering systems? Let us consider the excitation and emission processes in the pure and colloidal dye solutions. The key length scales are given by the inverse of the product of a cross-section and density. The cross-sections for l , l_a (the pump absorption length) and l_{sr} are for scattering, absorption and stimulated emission, respectively, and the corresponding densities are of scatterers, of the excess of dye molecules in the ground over the excited state for the pumped transition, and of the corresponding excess of the excited state over the ground state of molecules in the emitting transition.

In the low-power limit in the pure dye solution, the pump intensity decays exponentially on a scale, l_a , determined by the ground-state population density. As the power increases, absorption is suppressed as the excited-state population increases. This pulls the system towards super-radiance because on the one hand the excited state population of the emitting state is enhanced and on the other the absorption is reduced, leading to an ex-

tended excitation region.

In strongly scattering colloidal systems, the penetration depth of radiation in the limit of low incident power is reduced to a depth of $L_a = (l_a/3)^{1/2}$, where l_a is now the absorption length at a typical value of the spatially varying pump intensity. But as the pump power increases, l_a increases, leading to a far greater penetration depth. Just as in the pure dye solution, this enhances the power deposited in the sample and lengthens the typical

path length of emitted photons in the gain region. In the absence of gain or loss for the emitted photons, the typical path length of an emitted photon from a depth L_a would be approximately equal to the absorption length l_a of the exciting photon. But much longer path lengths may be anticipated because photons experience gain along the path, so that the weight of such paths in the distribution of emitted light is increased. So any means of increasing the path length of photons in the region overlapping the exciting radiation, such as internal reflection by a large index mismatch at the sample boundary or resonances within scattering spheres containing dye molecules, may aid lasing action.

If we conservatively estimate the scattering cross-section to be the cross-sectional area of the particle, we find l is about 20 μm at the highest particle concentration studied. For the system discussed here, the low power value of l_a is about 100 μm and l_{sr} is about 400 μm . Thus in the absence of saturation we would not expect to observe laser action. But once the threshold for saturation is reached, the factors given above work together to

stretch the typical path length of emission to a value beyond l_{sr} . The arguments given above require that l is smaller than l_a and L_a . In the opposite limit of weak scattering, the scatterers would destroy the coherence of the emission and suppress the super-radiance, as is indeed observed (Siddiq *et al.*, personal communication). Quite surprisingly, Lawandy *et al.* observe super-radiant emission induced by the presence of scatterers even in samples that are thinner than the scattering length. This may be the result of the trapping of a fraction of the scattered light by total internal reflection at the cuvette's internal surfaces. The phenomenon is reminiscent of the observation of super-radiance by R. L. Fork (*Phys. Rev. B* **19**, 3365–3398; 1979) in rare earth microcrystals.

In a related development, A. Yu. Zyuzin has predicted recently that the coherent backscattering from an amplifying slab of random scatterers should exhibit a narrow peak at a critical thickness at which the system is close to the super-radiant threshold (*Europhys. Lett.*, in the press). At this thickness the amplification of longer paths produces a uniform spatial profile of reflected light from point excitation instead of a narrow intensity distribution peaked about the point of incidence.

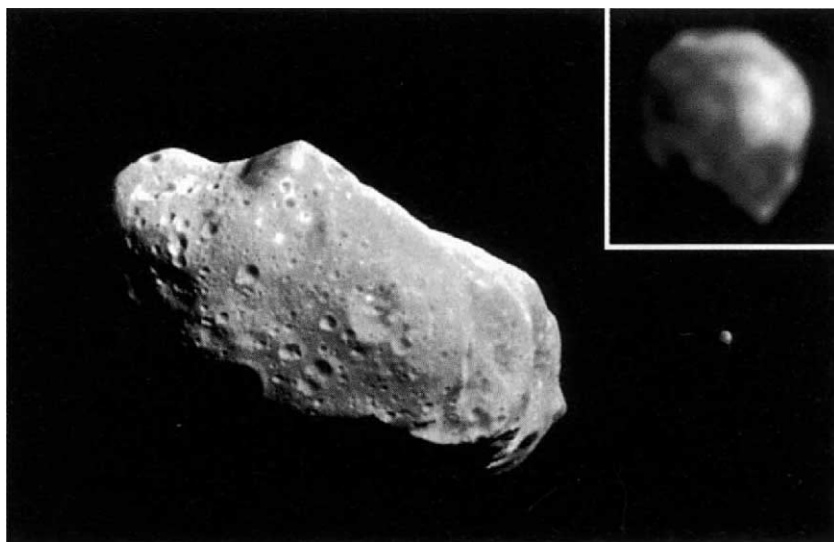
The observation of gain in random media opens up the possibility of developing compact super-radiant devices. Key experiments which would help establish the usefulness of such systems would be to measure the depth to which the pump radiation penetrates and the path length distribution of emitted light, which can be established using ultrashort optical pulses. □

A. Z. Genack is at Queens College of the City University of New York, Flushing, New York 11367, and J. M. Drake is at the Exxon Research and Engineering Company, Annandale, New Jersey 08801, USA.

ASTEROIDS

New moon

A CHIP off the old block? Photographs and spectrographic maps returned by the Galileo spacecraft, and released by NASA last week, show a tiny moon accompanying the asteroid 243 Ida; this is the first confirmed natural satellite of any asteroid. Ida herself is about 56 kilometres long; and her far smaller companion (also shown enlarged) may be a fragment of the same large parent asteroid, or the relic of a more recent blow to Ida. Galileo recorded the images last August as it flew through the asteroid belt more than 500 million kilometres from Earth. The moon, first spotted in a sample image strip in mid-February, should be revealed ever more clearly in the next few months as further images and spectra are returned. L. M.



NASA JPL

Simplifying the complex

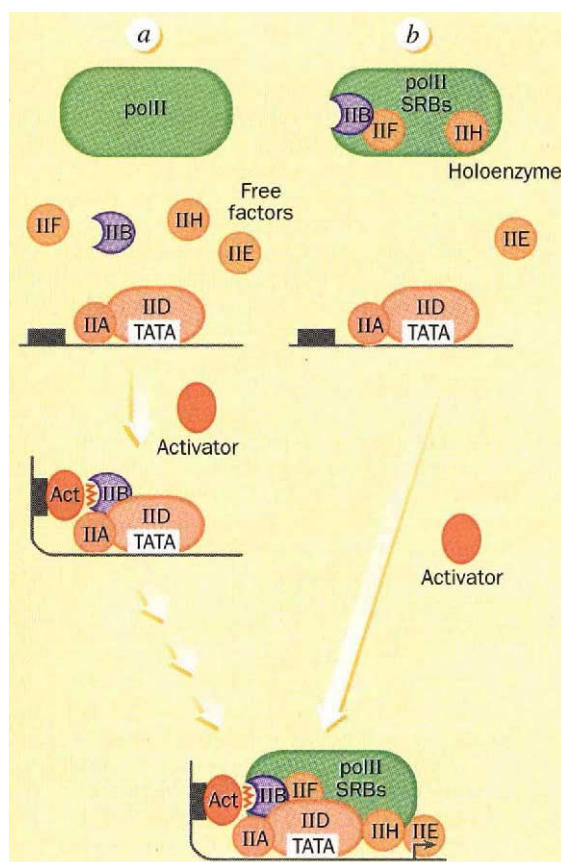
Michael Carey

OVER a decade ago, studies of gene expression established the concept that transcription in eukaryotic cells is mediated by a multifactor, nucleoprotein ensemble known as the transcription complex. The prevailing hypothesis has been that the complex assembles on promoter DNA from free factors in an ordered, stepwise fashion. This view now comes under challenge from Koleske and Young (page 466 of this issue¹), who report the purification of a largely preassembled RNA polymerase II (pol II) transcription complex from crude homogenates of *Saccharomyces cerevisiae*. The finding simplifies models for eukaryotic gene activation and indicates key regulatory checkpoints in the process.

The story is a classic example of molecular detective work, beginning in the mystical world of genetics and ending in the material world of biochemistry. Partial deletions in the carboxy-terminal domain of the pol II large subunit elicit a cold-sensitive (*cs*) growth phenotype in yeast. The carboxy-terminal domain is composed of 26 copies of a heptapeptide repeat, whose function in gene regulation has been the subject of considerable research and speculation. In a previous study², Young and colleagues isolated genetic suppressors of the *cs* mutants, called SRBs (suppressor of RNA polymerase B), and found that the SRB-encoded proteins copurified in a stable complex with pol II. Although the complex contained only 6 per cent of the cellular pol II, it comprised 65 per cent of the SRBs.

Koleske and Young now report that the complex also contains a subset of the yeast general transcription factors^{3,4}. In many eukaryotic organisms, several general factors (TFIIA – TFIIF) are required to allow pol II to initiate specific transcription at a core promoter (that is, TATA box and startsite)^{3,4}. The best characterized of them is TATA-binding protein (TBP), which in *Drosophila* and mammalian cells is tightly complexed with TBP-associated factors (TAFs) that, in part, mediate the action of upstream activators; the TBP–TAF complex is called TFIID. The current view, based on an abundance of biochemical data^{3,4}, is that preinitiation complex assembly *in vitro* is nucleated by binding of TBP to the TATA box and is followed by the stepwise, ordered addi-

tion of the remaining factors and pol II. It therefore came as a surprise that the SRB complex identified by Young and co-workers could mediate both basal and GAL4-VP16-activated transcription upon addition of only TBP and TFIIE. This apparent paradox was solved when fur-



Models for the assembly of the transcription complex by activators. *a*, The multistep pathway. TATA-binding protein (complexed with TAFs to form TFIID) binds to the TATA box, TFIID binding being stabilized by TFIIA. The activator binds its recognition site in the DNA and recruits TFIIB to the complex; the remaining factors and polymerase II are added in steps. *b*, The simplified pathway. TFIID/TFIIA binds the TATA box; the activator then binds and recruits the holoenzyme and TFIIE.

ther analysis of the purified complex revealed that it contained stoichiometric quantities of TFIIB, TFIIF and TFIH. The multifactor nature of the SRB complex inspired Koleske and Young to name it the 'holoenzyme'.

It is unclear why the holoenzyme had not been previously observed. Perhaps it is because the complex contains only a small proportion of total cellular pol II or general factors. Alternatively, the integrity of the complex may depend upon the method of fractionation. Either way, given its novelty, the relevance of the discovery will probably be called into

question. The most cynical view is that the holoenzyme represents the elongating form of pol II. The main argument against this point is that the pol II elongation complex apparently does not contain TFIIB (ref. 5), while the holoenzyme does. Furthermore, the general factors found associated with the holoenzyme are known to bind RNA polymerase in solution, so it is not unexpected that they should exist in a complex⁴. Finally, it must be emphasized that our current under-

standing of eukaryotic transcription initiation is derived from a strictly biochemical approach without the foresight of genetics, whereas the holoenzyme is the biochemical manifestation of a genetic phenotype, supporting its relevance *in vivo*.

Will the holoenzyme radically alter our view of eukaryotic transcription? Probably not, but it represents a noteworthy revision. The attractive aspect of the discovery is that it simplifies ideas about regulation. It might be imagined that activators, rather than affecting myriad steps in preinitiation complex assembly, will recognize one of many available surfaces on a single contiguous holoenzyme 'target' and, in cooperation with TFIID and TFIIE, recruit it to the promoter. The chimaeric activator GAL4-VP16, which has previously been shown to interact tightly with TFIIB (ref. 6) *in vitro*, can now recruit the entire complex in one fell swoop (see figure). Notably absent from the holoenzyme is TFIIA, which is not required for basal or activated transcription in Koleske and Young's experiments. Our current view of TFIIA is that it stabilizes TFIID binding to TATA and competes with negative cofactors that adhere to TFIID and block assembly of the preinitiation complex⁴. So TFIIA would only be required in crude systems, or in the cell where negative factors such as the recently described ATP-dependent inhibitor (ADI)⁷ might be present.

The most exciting prospect is that the holoenzyme will chart new directions in which to approach eukaryotic transcription. Because many polypeptides in the complex are undefined, it may serve as a source for pleiotropic regulatory mol-

1. Koleske, A. J. & Young, R. A. *Nature* **368**, 466–469 (1994).
2. Thompson, C. M. *et al.* *Cell* **69**, 883–894 (1992).
3. Zawal, L. & Reinberg, D. *Prog. Nucleic Acids Res.* **44**, 67–108 (1993).
4. Conaway, R. C. & Conaway, J. W. *A. Rev. Biochem.* **62**, 161–190 (1993).
5. Reines, D. *J. Biol. Chem.* **266**, 10510–10517 (1991).
6. Choy, B. & Green, M. R. *Nature* **366**, 531–535 (1993).
7. Auble, T. A. & Hahn, S. *Genes Dev.* **7**, 844–856 (1993).

ecules identified in other yeast genetic screens. The complex may also provide insight into the action of TAFs, because two of the SRBs have TBP-binding activity^{1,2}. Finally, analysis of the holoenzyme should help resolve the controversial role in pol II initiation of phosphorylation of the carboxy-terminal domain^{3,4}. Such questions aside, the most pressing issue is whether the stepwise or

holoenzyme pathways (or both) are operational *in vivo* — that's a daunting experimental problem, but one which is central for understanding eukaryotic gene regulation. □

Michael Carey is in the Department of Biological Chemistry, UCLA School of Medicine, University of California, Los Angeles, California 90024, USA.

SPIRAL GALAXIES

A dynamo with a difference

Katia Ferrière

THE existence of magnetic fields in spiral galaxies has been established for over three decades, but their origin is still a matter of controversy. The most widely accepted theory is that they arise through the action of a hydromagnetic dynamo, but despite its successes at reproducing observed magnetic fields, the standard form of this theory encounters difficulties which have led several researchers to look to complementary or alternative mechanisms. One compelling alternative is put forward by Chakrabarti, Rosner and Vainshtein on page 434 of this issue¹.

The first step is to recognize that galactic magnetic fields are unlikely to be of primordial origin, in other words the relic of a magnetic field existing before the galaxy formed². If a coherent magnetic field indeed existed in the early Universe, the gas motions associated with the collapse of a galaxy would have compressed the lines of force of the ambient magnetic field (field lines tend to be frozen into the highly conductive gas), and from then on, the differential rotation of the galaxy would have wrapped them up about its centre; by now, the magnetic field would be wound up roughly 50 times, in disagreement with observations which never reveal tightly wound galactic field configurations.

The traditional way of resolving the discrepancy is to appeal to magnetic diffusion (presumably of the turbulent kind). But if the galactic magnetic field diffuses horizontally fast enough to avoid substantial wind-up, it also diffuses vertically out of the galactic disk in a time shorter than the rotation period, so it should have decayed away to nothing by now (this conclusion possibly no longer holds if one appeals to ambipolar diffusion, which allows magnetic fields, tied to the charged particles, to slip through the neutral medium³). Hence, it is usually believed that galactic magnetic fields can be explained only if there exists a mechanism to regenerate them.

In the dynamo theory, fields are regenerated through the action of small-scale turbulent motions, termed the

'alpha-effect'⁴. Qualitatively this can be described as follows. Because turbulent motions take place in a rotating medium, they are cyclonic; that is, they have a preferred sense of rotation. As they stretch and twist magnetic field lines, they impart to them a net rotation, whereby a magnetic field can be created in the direction perpendicular to the prevailing field. It is the combination of the large-scale differential rotation and small-scale cyclonic turbulence that leads to magnetic field amplification: the differential rotation stretches field lines in the azimuthal (longitudinal) direction about the galactic centre, while the alpha-effect regenerates the meridional component of the field from its azimuthal component.

Obviously a weak initial magnetic field is necessary to trigger the dynamo process; this seed field could have a pregalactic origin (as described above), be produced by a battery effect due to charge separation in the protogalaxy (galaxy precursor), or else be injected into the interstellar medium by the first generation of stars⁵.

At present, it is impossible to solve the galactic dynamo problem by direct numerical simulations without making some assumptions about the properties of turbulent processes, because even the most powerful computers are unable to cover the whole range from large galactic scales down to the smallest turbulent scales.

The standard approach is to assume that the effects of turbulent motions on the large-scale magnetic field, **B**, are entirely contained within two parameters (hereafter referred to as the dynamo parameters): the alpha-coefficient, which embodies the alpha-effect, and the turbulent magnetic diffusivity⁶. In the absence of more detailed understanding, one must usually use order-of-magnitude estimates for the dynamo parameters, for instance based on mixing-length theory or on crude pictures of interstellar turbulence. A further simplification is to consider the galactic rotation and the dynamo parameters to be independent of the magnetic

field, in which case the evolution of **B** is governed by a linear equation whose solutions grow (or decay) exponentially with time.

But as soon as magnetic energy becomes significant compared to kinetic energy — the system approaches equipartition, to put it another way — the back-reaction of the generated magnetic field on the gas motions can no longer be ignored. The Lorentz force suppresses turbulence, thereby reducing the dynamo parameters. The large-scale differential rotation is also affected, although to a lesser extent because its kinetic energy is at least two orders of magnitude larger than the turbulent kinetic energy.

Recent galactic dynamo calculations⁷ have attempted to take the magnetic feedback into account by letting the dynamo parameters be decreasing functions of the large-scale magnetic field strength; in this case, the evolution equation for **B** is nonlinear, and its growing solution saturates when the large-scale magnetic energy comes into equipartition with the turbulent kinetic energy. The nonlinear calculations are certainly more realistic than their early kinematic counterparts. But they face the same fundamental problem: they represent the complex effects of turbulence by only two parameters, whose values are no more than a rough guess and whose functional forms are chosen for mathematical convenience rather than on a sound physical basis.

According to Vainshtein and Cattaneo⁸, the influence of turbulent processes on the large-scale magnetic field is overestimated by the current dynamo theory. Their argument goes as follows: field generation through the alpha-effect and field destruction through turbulent diffusion are possible only if the magnetic field is not perfectly frozen into the gas, and if turbulent magnetic structures are formed down to the very small scales at which molecular processes lead to a violation of the frozen-in condition, allowing the magnetic field to diffuse and reconnect. The field strength at these very small diffusive scales is equal to the large-scale field strength times a power ($\geq 1/2$) of the magnetic Reynolds number, which, in galaxies, is huge. When the magnetic field at the diffusive scales reaches equipartition and saturates, the large-scale field is still far below equipartition. Hence, a self-consistent nonlinear dynamo theory cannot by itself explain why large-scale galactic magnetic fields are observed to be close to equipartition.

The alternative scheme now proposed by Chakrabarti *et al.*¹ also resorts to dynamo action, but only at an early stage of the generation process, which involves weak fields and is therefore far from saturation. Chakrabarti *et al.* proceed from the assumption that the centre of a young galaxy (still at the protogalactic

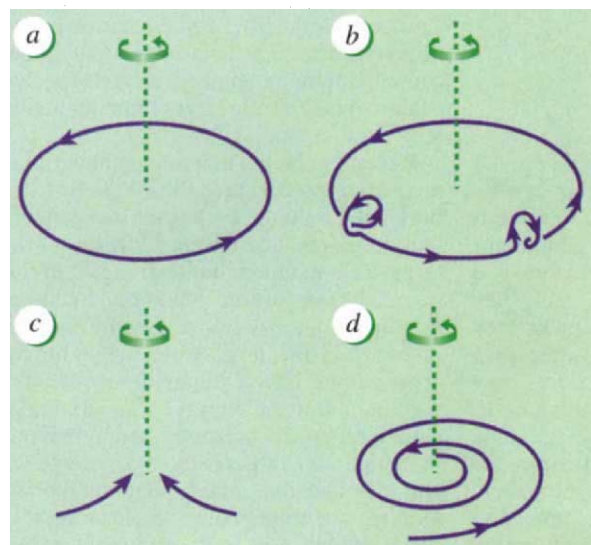
stage) contains a massive black hole as well as a thick plasma torus rotating about the black hole and at least partly supported by radiation pressure. Thermal battery processes in this torus create a weak azimuthal magnetic field, which, through the alpha-effect, gives rise to a weak meridional magnetic field (see figure). The meridional field is then sheared out in the azimuthal direction by the differential rotation, with the shear

black hole, whereas the classical dynamo would not have had enough time to amplify a seed magnetic field up to the observed values.

Other implications of the model should also be confronted with observations. In particular, do black holes systematically form early in the cores of galaxies? And is the predicted relationship between the black hole mass and the magnetic field strength actually observed?

In brief, explaining the origin of galactic magnetic fields remains a challenge. The well-developed dynamo theory offers a physically compelling scenario, and appears to give good results in that it manages to reproduce the observed large-scale field structures. But it dodges the crucial question of exactly how small-scale turbulent motions interact with the large-scale magnetic field that they are supposed to amplify. If, for instance, magnetic suppression of the alpha-effect and of the turbulent diffusion is as severe as claimed by Vainshtein and Cattaneo⁸, another mechanism clearly must come into play.

The idea that magnetic fields are first generated near the galactic core, where the local differential rotation stretches field lines very efficiently, and are then transported outwards



Galactic magnetic field generation according to Chakrabarti *et al.*¹. a, A weak azimuthal magnetic field is produced by a thermal battery in the plasma torus around the massive black hole assumed to lie at the heart of the galaxy; b, this azimuthal magnetic field is stretched and twisted by cyclonic turbulent motions (alpha-effect) to produce a meridional magnetic field as shown in c; d, the meridional magnetic field is stretched in the azimuthal direction by differential rotation.

near the black hole being so strong that the resulting azimuthal field reaches local equipartition in less than a thousand years or so. Finally, the magnetic field produced near the galactic centre is rapidly distributed throughout the disk via a wind or some other kind of outflow.

How well does this square with observations? The new theory predicts that the large-scale azimuthal magnetic field falls off in inverse proportion to the galactocentric radius, R . Under these conditions and for realistic values of the central black hole mass, it is then possible to obtain the observed magnetic field strength near the Sun. Unfortunately, the predicted R^{-1} dependence of the galactic azimuthal field is not supported by observations⁹; in this respect, the magnetic field configurations obtained in the standard dynamo theory agree better.

On the other hand, the Faraday rotation detected in the absorption clouds of some quasars suggests that strong magnetic fields exist in very young galaxies^{10,11}. If confirmed, this fact would be consistent with Chakrabarti and co-workers' model, which allows for rapid magnetic field growth after the formation of the central

by a wind, is certainly appealing. It solves a number of problems inherent in the standard dynamo theory, but it raises its own set of questions. Its fate will ultimately depend on how well these questions can be answered. □

Katia Ferrière is in the Observatoire Midi-Pyrénées, 14 avenue Edouard-Belin, F31400 Toulouse, France.

1. Chakrabarti, S. K., Rosner, R. & Vainshtein, S. I. *Nature* **368**, 434–436 (1994).
2. Rosner, R. & DeLuca, E. in *The Center of the Galaxy*, IAU Symp. 136 (ed. Morris, M.) 319–328 (Kluwer, Dordrecht, 1989).
3. Kulsrud, R. In *Galactic and Extragalactic Magnetic Fields*, IAU Symp. 140 (eds Beck, R., Kronberg, P. P. & Wiebeinski, R.) 527–530 (Reidel, Dordrecht, 1990).
4. Parker, E. N. *Astrophys. J.* **163**, 255–278 (1971).
5. Rees, M. J. *Q. J. R. astr. Soc.* **28**, 197–206 (1987).
6. Moffatt, H. K. in *Magnetic Field Generation in Electrically Conducting Fluids* (eds Batchelor, G. K. & Miles, J. W.) 145–178 (Cambridge Univ. Press, 1978).
7. Brandenburg, A. *et al. Astr. Astrophys.* **259**, 453–461 (1992).
8. Vainshtein, S. I. & Cattaneo, F. *Astrophys. J.* **393**, 165–171 (1992).
9. Heiles, C. in *Interstellar Processes* (eds Hollenbach, D. J. & Thronson, H. A.) 171–194 (Reidel, Dordrecht, 1987).
10. Kronberg, P. P. & Perry, J. J. *Astrophys. J.* **263**, 518–532 (1982).
11. Wolfe, A. M. in *QSO Absorption Lines: Probing the Universe* (eds Blades, J. C., Turnshek, D. & Norman, C. A.) 297–317 (Cambridge Univ. Press, 1988).

DAEDALUS

Expanded metal

LAST week Daedalus pointed out that a metal consists of an array of positive ions neutralized by immersion in a sea of delocalized free electrons. He proposed to explore a whole range of new 'hybrid alloys', with organic ions nestling among the metallic ones.

A metal-rich alloy, with only a small number of included alien ions, would simply be a modified metal. A roughly equal number of metallic and non-metallic ions might pack together badly. A metal-poor alloy would again pack well, this time with metal ions in the interstices of the lattice. If he can make an alloy of this composition, Daedalus wants to heat it. He recalls the trick of heating solvated silica gel above the critical point of the included solvent. The vapour escapes, leaving an amazingly light, open silica framework — an aerogel. Silica aerogel can be almost as light as air: it looks like (and is known as) 'frozen smoke'. Daedalus reckons that on heating a metal-poor hybrid alloy, the major organic component would similarly vaporize, leaving a sparse distribution of metal atoms as a metallic aerogel.

Metallic aerogel will be uncannily light and tenuous, yet utterly black. It should conduct electricity fairly well, but heat badly. Immersed in plastic monomer or molten glass, it could make novel metal-reinforced composites. Unlike other aerogels, it should be ductile. It could be formed into ghostly black wire, sheet and engineering parts of great toughness yet weighing almost nothing.

The ultimate metal-poor hybrid alloy would have no metal at all. It would consist entirely of non-metallic positive ions in a sea of free electrons.

Ammonium (NH_4^+), or its tetramethyl derivative, seem the most promising candidates; free electrons dissolve in liquid ammonia, and neutral ammonium forms an amalgam with mercury. Atoms contract on forming positive ions, so metallic ammonium should be made by compressing ammonia with hydrogen.

Once made, metallic ammonium should be stabilized by electronic delocalization, and might persist even at atmospheric pressure. It would probably resemble lithium or sodium: very light but rather soft, too reactive to be much use structurally, but valuable in alloys and for batteries. It might, however, be rather explosive, reverting violently to ammonia and hydrogen gas: it could make a splendid rocket fuel. But nature may have got there first. Daedalus suspects that Jupiter, which contains abundant ammonia, hydrogen and methane under great pressure, is largely made of metallic ammonium or tetramethyl ammonium. David Jones

Degassing of Lake Nyos

SIR — Lakes with CO₂-rich waters are distinguished by their extraordinary chemical evolution, their lethal nature and their uniqueness among natural hazards in potential for mitigation. Three possible endpoints of chemical evolution exist for these systems: equilibrium, where gas input is balanced by gas loss; gas bursts, where recharge surpasses loss; and controlled degassing, where the cycle of recharge and release is short-circuited. Here we report the use of direct measurements of CO₂ recharge rates and introduce a new parameter of lake stability to evaluate the likelihood of equilibrium or of further gas bursts, and to analyse the potential of controlled degassing.

The changing inventory of CO₂ in these lakes is the difference between inputs of dissolved gas issuing from bottom or side vents¹⁻⁴, and losses including leakage, surface flushing and ventilation to the atmosphere. Since the 1986 gas burst in Lake Nyos, Cameroon, a total of 19×10^8 mol CO₂ have been lost through flushing and ventilation as the surface mixing layer deepened to 50 m, whereas gas inputs resulted in an accumulation of 11×10^8 mol CO₂ below 50 m (2×10^8 mol yr⁻¹;

table). It is unlikely that losses from surface waters will continue to be a significant counterbalance for gas recharge. The surface layer is now chemically similar to the surface layer before the event², and data on controls of mixing depth in tropical lakes⁵ suggest that strong density gradients at 50 m in Nyos will inhibit further substantial deepening. Diffusional loss of CO₂ from below 50 m into the surface layer will continue; the current rate is about 0.17×10^8 mol yr⁻¹, or 8% of recharge rate.

In Lake Monoun, Cameroon, the surface mixing layer is shallow (5 m), and has changed little since 1987. Consequently, gas inputs were weakly compensated by surface losses, and the total CO₂ inventory increased from 5.41×10^8 to 6.24×10^8 mol (0.17×10^8 mol yr⁻¹). Direct measurements of gas recharge in both lakes confirm the hypothesis of slow accumulation and gas storage in bottom waters^{1,6} rather than rapid injection of gas by phreatic explosions⁷, and highlight the continuing danger of violent degassing.

Gas release is controlled by local stability and the ratio of total gas pressure (P_{gas}) to ambient hydrostatic pressure (P_{amb}) (Fig. 1). Dissolved CO₂ has a non-linear effect on stability, increasing water density and stability until the ratio $P_{\text{gas}}/P_{\text{amb}}$ exceeds 1, at which point turbulence from exsolving gas destroys local stability, and possibly triggers a gas burst. The maximum $P_{\text{gas}}/P_{\text{amb}}$ is 0.55 in Nyos (at 209 m) and 0.78 in Monoun (99 m) (Fig. 1). If current rates of pressure buildup continue, $P_{\text{gas}}/P_{\text{amb}}$ will reach 1 in 30 ± 8 yr in Nyos and in only 7 ± 2 yr in Monoun.

Knowledge of local stability is required to model the process of gas release, to predict hazard potential, and to determine the optimal procedure for controlled degassing. Because standard measures of local stability fail to account for the effects of dissolved gas, we introduce a modified stability parameter

$$E^* = [(1/\rho)(dp/dZ)] \times [(P_{\text{amb}}/P_{\text{gas}}) - 1] \times (1/P_{\text{gas}})$$

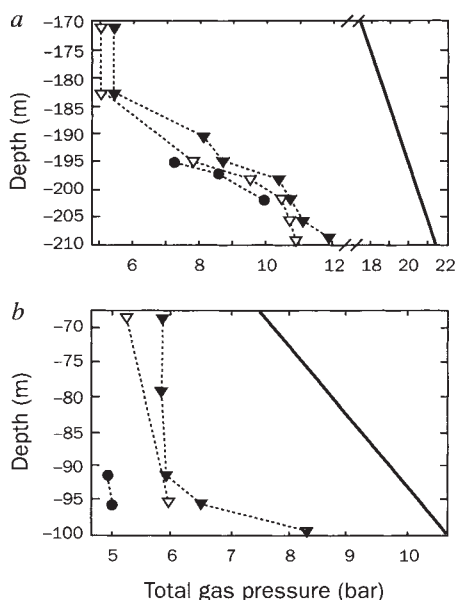


FIG. 1 Total gas pressure measured *in situ* by probe in Lakes Nyos (a) and Monoun (b) (total probe error, ± 0.15 bar²). P_{gas} at 171 and 183 m in Nyos was calculated as the sum of P_{CO_2} , P_{CH_4} and P_{N_2} (99% of total P_{gas}) measured in samples from pressurized cylinders. Average changes in P_{gas} were calculated for each depth and between each date, and then compared to the 1992 gas pressures to derive the time required for P_{gas} to reach P_{amb} (solid line). All rates and times were then statistically averaged; there was no statistically significant difference in buildup rate in 1989–90 compared to 1990–92. Filled circles, Dec 1989; open triangles, Sep 1990; filled triangles, Mar 92.

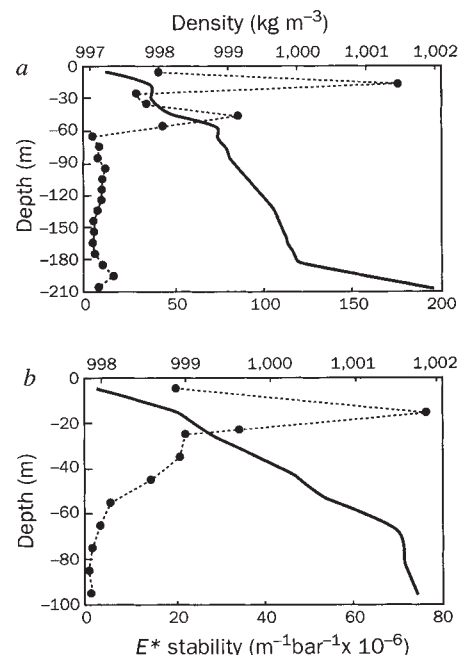


FIG. 2 Potential density (ρ ; solid line) and local water column stability including the effects of dissolved gas (E^* ; dotted line) in Lakes Nyos (a) and Monoun (b). Final potential density due to temperature, dissolved salts, and gases was calculated as in ref. 4. Bathymetric and dissolved salt data are from refs 2,4; dissolved salt data from 1992 are unpublished, but are similar to September 1990 data². The first term of E^* , $(1/\rho)(dp/dZ)$, is a standard measure of stability; the second term $(P_{\text{amb}}/P_{\text{gas}}) - 1$, describes the pressure ratio such that as P_{gas} approaches P_{amb} the stability E^* goes to zero; the final term, $(1/P_{\text{gas}})$, is required to distinguish situations where local saturation is reached in a layer with low absolute gas pressure, such as near the lake surface, and thus a low potential for triggering a gas burst. Negative values of E^* indicate unstable conditions due to denser water overlying less dense water (first term of E^*) or to gas oversaturation (second term of E^*).

where ρ is potential density of the fluid and Z is depth (Fig. 2). Accordingly, as P_{gas} rises the local stability E^* is reduced.

It has been suggested that degassing the uppermost CO₂-rich layers is a priority during mitigation⁸; we disagree. During controlled degassing, water pumped from a specified depth and degassed at the surface may be released in the lake, or it may be removed from the basin so that lake level is lowered. In the first case, E^* increases everywhere above the pumping depth and remains constant elsewhere. In the second, E^* decreases at all points below the degassing depth and remains constant elsewhere; stability decreases because lowering lake level decreases P_{amb} while P_{gas} remains constant. Therefore, regardless of the degassing procedure, the safest and most stable conditions will be attained when pumping occurs from the deepest zones of low E^* . Zones of low E^* are also the most likely initiation points in

- Kling, G. W. *et al.* *Science* **236**, 169–175 (1987).
- Evans, W. C., Kling, G. W., Tuttle, M. L., Tanyileke, G. & White, L. D. *Appl. Geochem.* **8**, 207–221 (1993).
- Nojiri, Y. *et al.* *Nature* **346**, 322–323 (1990).
- Kling, G. W., Tuttle, M. L. & Evans, W. C. *J. Volcanol. geotherm. Res.* **39**, 151–165 (1989).
- Kling, G. W. *Limnol. Oceanogr.* **33**, 27–40 (1988).
- Sigurdsson, H. *et al.* *J. Volcan. geotherm. Res.* **31**, 1–16 (1987).
- Tazieff, H. *J. Volcan. geotherm. Res.* **39**, 109–116 (1989).
- Fletcher, H. *Nature* **355**, 683 (1992).
- Kling, G. W., Tuttle, M. L. & Evans, W. C. *Nature* **337**, 215 (1989).
- Freeth, S. J. *et al.* *Nature* **348**, 201 (1990).
- Halbwachs, M., Grangeon, J., Sabroux, J.-C. & Villavieille, A. C. R. *Acad. Sci. Paris* **316**, 483–489 (1993).
- Lockwood, J. P. & Schuster, R. L. *Nature* **351**, 195 (1991).
- Kling, G. W., Evans, W. C. & Tuttle, M. L. *Verh. int. Verein. Limnol.* **24**, 1102–1105 (1991).

Depth CO ₂ (m) (μmol kg ⁻¹)	Depth CO ₂
LAKE MONOUN	LAKE NYOS
Dec 1989	Apr 1992
23 17,900	53 55,500
46 54,300	76 71,800
69 145,000	99 92,500
95 120,000	137 130,000
79 138,000	171 144,000
95 140,000	183 155,000
Sep 1990	190 209,000
23 20,400	198 289,000
46 64,600	206 318,000
61 119,000	206 319,000
69 140,000	209 315,000
Mar 1992	
23 21,100	
69 145,000	
79 14,800	
95 150,000	

Concentrations of dissolved gas in Lakes Nyos and Monoun measured in pre-evacuated cylinders². Recharge rates are calculated using earlier data for Nyos^{2,4} and Monoun¹³.

triggering a gas burst.

Dangerous gas-rich lakes can be identified before gas bursts occur³, and alleviation of the hazards is possible through

controlled pumping in pipes^{1,10,11}. Preliminary calculations based on field tests¹¹ and our measured inventory and recharge indicate that Monoun could be fully degassed in less than 2 yr using three pipes of 141-mm diameter. The 550,000 tonnes of CO₂ in Lake Nyos will require more or larger pipes and a longer period of time, and complications arise due to the geologically weak spillway of the lake¹². Permanent installation of the pipes would short-circuit the natural cycle of gas recharge followed by gas burst in these lakes. Controlled degassing of the lakes is needed urgently.

George W. Kling
William C. Evans
Michele L. Tuttle
Greg Tanyileke
Department of Biology,
University of Michigan,
Ann Arbor,
Michigan 48109, USA

Other addresses: US Geological Survey, Menlo Park, California 94025, USA (W.C.E.); US Geological Survey, Denver Federal Center, Denver, Colorado 80225, USA (M.L.T.); Institute for Geological and Mining Research, Mesres, Yaounde, Cameroon (G.T.).

Here we demonstrate the preferential migration of implanted mononuclear myoblasts from isografts to regions of muscle injury and regeneration in adjacent muscles.

We injected mouse C2C12 myoblasts carrying the *lac Z* gene¹¹ encoding β-galactosidase (β-gal) into a mouse extensor digitorum longus muscle (EDL) subsequently isografted into a second host. All recipients were athymic nude mice bearing the mouse X-linked muscular dystrophic (*mdx*) gene and therefore lacking dystrophin¹². Spongostan provides a penetrable support medium¹³, and we inserted it between the isograft and surrounding muscles. We injured tibialis anterior (TA) muscles and/or peroneal muscles using forceps.

The presence of β-gal- and dystrophin-positive fibres in the EDL isografts confirmed the incorporation of implanted cells (Fig. 1a, b). Injury to one or more neighbouring host muscles resulted in β-gal- and dystrophin-positive myotubes and fibres in the Spongostan adjacent to the injured muscle (Fig. 2a, b). In the case of single muscle injury (9 cases), we saw positive fibres in the adjacent injured muscle but not in the neighbouring uninjured muscle (Fig. 2c, d). Where two host muscles were injured (2 cases), implanted cells migrated to both (Fig. 2e). These results indicate the preferential migration of implanted cells from the isograft, through the Spongostan to a site of muscle injury. Where host muscles were uninjured, there was no evidence of migration of implanted cells, either in the Spongostan or adjacent muscles in two of the three cases. But in the third case, we saw β-gal/dystrophin-positive fibres in both TA and peroneus (Fig. 2f). This finding

Migration of muscle cells

SIR — Immature muscle cells (myoblasts) can deliver gene products to mature muscle fibres following intramuscular or systemic delivery^{1–4} and thus could have a therapeutic role in inherited muscle diseases^{1,5}. Successful therapy would re-

quire migration of introduced cells throughout the muscle so that gene products are disseminated diffusely. Myoblast migration is controversial, many reports showing movement within muscles, and few movement between them^{6–10}.

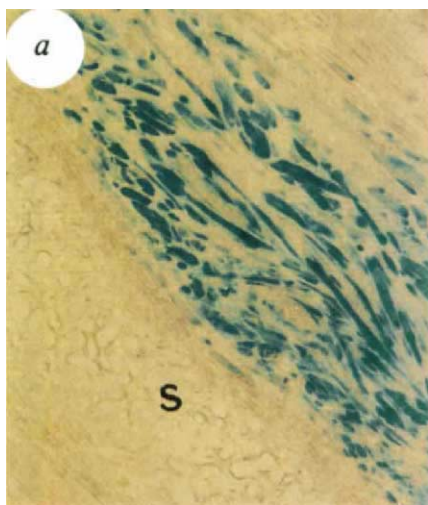


FIG. 1 Micrographs showing the EDL isograft 17 days after implantation. Most EDL fibres are positive for (a) β-gal (blue) X 180, and (b) dystrophin, X 270, S; Spongostan.

METHODS: Isografting was performed by removal of C2C12-implanted EDL from the *mdx nu/nu* host and transplanted into the EDL muscle bed of a second *mdx* recipient from which the autologous EDL had been removed. Using a PCR pipette flamed to a fine point, the EDL to be isografted was injected with 3×10^5 C2C12-*lacZ* cells 3–5 days before their removal from the initial host. Muscle complexes removed 15, 17 or 22 days following isograft insertion and crush injury, were frozen in isopentane maintained at -165°C in liquid nitrogen. Cryostat sections ($8\mu\text{m}$) were cut at three levels throughout the complex for examination by light and electron microscopy. Analysis for β-gal activity was carried out as described¹⁶, but leaving the reaction product to develop overnight. For dystrophin analysis¹², sections were immunocytochemically stained using an anti-dystrophin antibody of *M_r* 60,000, a gift from E.P. Hoffman.

1. Partridge, T.A., Morgan, J. E., Coulton, G. R., Hoffman, E. P. & Kunkel, L. M. *Nature* **273**, 176–179 (1989).
2. Neumeyer, A. M., DiGregorio, D. M. & Brown, R. H. *Neurology* **42**, 2258–2262 (1992).
3. Barr, E., & Leiden, J. M. *Science* **254**, 1507–1512 (1991).
4. Dhawan, J. *et al. Science* **254**, 1509–1521 (1991).
5. Watt, D. J., Morgan, J. E. & Partridge, T. A. *Muscle & Nerve* **7**, 741–750 (1984).
6. Lipton, B. H. & Schultz, E. *Science* **205**, 1292–1294 (1979).
7. Ghins, E., Colson-Van Schoor, M., Maldague, P. & Marechal, G. *Archs Int. Phys. Biochim.* **93**, 143–153 (1985).
8. Hughes, S. M. & Blau, H. M. *Nature* **345**, 350–353 (1990).
9. Phillips, G. D., Hoffman, J. R. & Knighton, D. R. *Cell Tissue Res.* **262**, 81–88 (1990).
10. Watt, D. J., Karasinski, J. & England, M. A. *J. Mus. Res. Cell Motil.* **14**, 121–132 (1993).
11. Price, J., Turner, D. & Cepko, C. *Proc. natn. Acad. Sci. U.S.A.* **51**, 919–928 (1987).
12. Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* **51**, 919–928 (1987).
13. Gundersen, J. *Folia Angiol.* **28**, 121–122 (1980).
14. Venkatasubramanian, K. & Solrusch, M. *Dev Biol.* **104**, 428–433 (1984).
15. Robertson, T. A., Maley, M. A. L., Grounds, M. D. & Papadimitriou, J. M. *Exp Cell Res.* **207**, 321–331 (1993).
16. Dannenberg, A. M. & Suga, M. in *Methods, for Studying Mononuclear Phagocytes* (eds Adams, S. D. O., Edelson, P. J. & Koren, M. S.) 375–396 (Academic, New York, 1981).

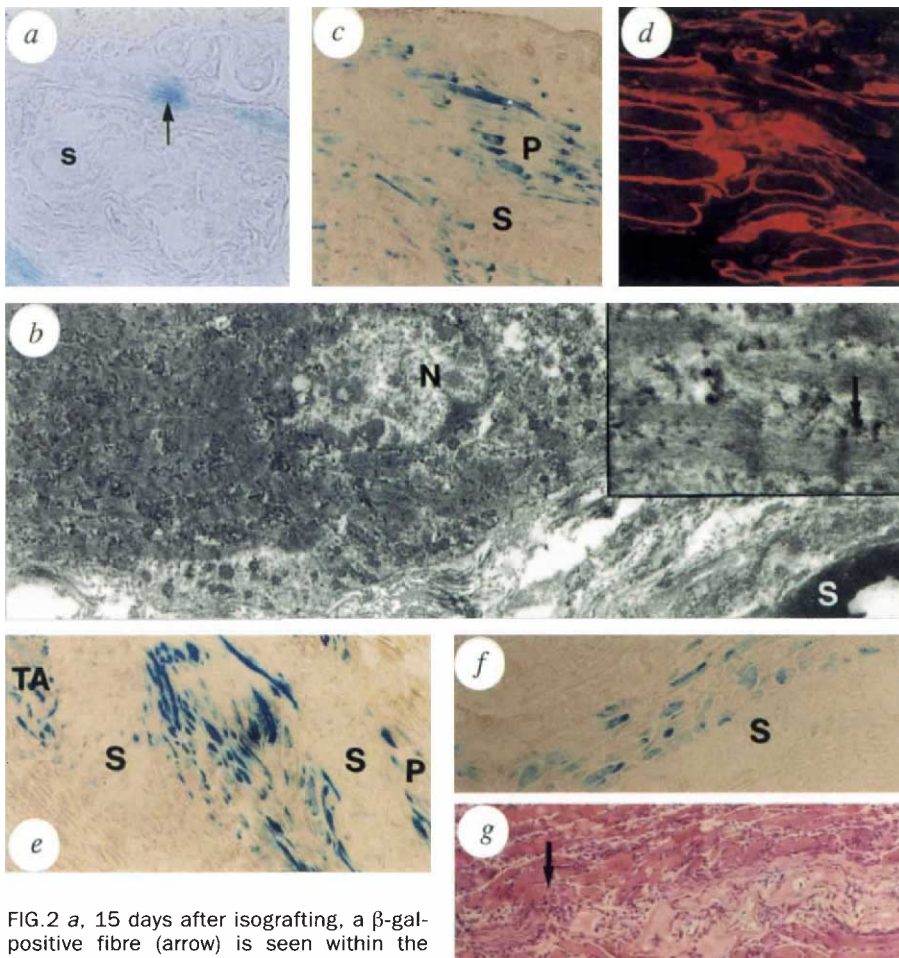


FIG. 2 *a*, 15 days after isografting, a β -gal-positive fibre (arrow) is seen within the Spongostan (S) between the isografted and injured TA, X 729. *b*, Electron microscopy was used to confirm that β -gal-positive cells within the Spongostan were myogenic and not macrophages (which endogenously express β -gal). EM of a cryostat section at 17 days post-isografting showing a myotube within the Spongostan between the isograft and injured TA. (N; nucleus of myotube) X 3,307. The β -gal reaction product is seen as granular electron dense material (arrow) in the sarcoplasm (inset X 12,150). *c*, Numerous β -gal-positive fibres throughout an injured peroneus (P), X 72.9. Fibres within this muscle also expressed the dystrophin gene product (*d*), X 364. *e*, β -gal-positive fibres in the isograft, TA and peroneus when both muscles were injured, X 72.9. *f*, β -gal-positive fibres in an uninjured, but spontaneously regenerating, peroneus, X 182. By haematoxylin and eosin staining the corresponding area (*g*) is seen to consist of narrow diameter regenerating fibres often with centrally located nuclei (arrowed, X 182).

correlates with the recent spontaneous regeneration, typical of *mdx*, seen in these muscles (Fig. 2g).

These results show that implanted cells migrate not only towards areas of muscle injury but also to regions of regeneration. Although the muscle into which the cells were initially implanted was deficient in the introduced gene products, it surrendered a proportion of implanted cells that passed through the Spongostan and continued their migration into the injured/regenerating muscles. This movement may be the result of a chemotactic influence exerted by injured fibres. Chemotactic behaviour of myoblasts has been previously reported¹⁴ and a recent study indicates that macrophages can exert such an effect on myogenic cells¹⁵.

In relation to human muscle diseases, our results indicate the importance of implanting cells when recipient muscle is regenerating to capitalize on the inherent

ability of such muscle to attract and incorporate the cells into its fibres. In terms of therapy, the results of these short-term studies using C2C12 cells are encouraging and provide a basis for further study using primary myoblasts.

D. J. Watt

J. Karasinski

J. Moss

*Departments of Anatomy
and Histopathology,
Charing Cross & Westminster
Medical School,
London W6 8RF, UK*

M. A. England

*Department of Anatomy,
University of Leicester,
Leicester LE1 9HN, UK*

*J. Karasinski is also at Department
of Cytology & Histology, Institute of
Zoology, Jagiellonian University, Krakow,
Poland.*

Punctuated debate

SIR — The review "Punctuated Equilibrium Comes of Age" by Gould and Eldredge¹ might lead the reader to suppose that the idea of evolution happening in fits and starts, and the evidence for it, were due to Gould and Eldredge. I remind them of the conference on genetics, palaeontology and evolution, published in 1949 and reprinted as a paperback in 1963 (ref. 2). A large part of this book, whose authors include many of the leading palaeontologists of the day, is about storms and doldrums in evolution; indeed, in summarising the conference, H. J. Muller writes "There have been unquestionable spurts and bursts, and contrasted long periods of relative stasis". All this was written more than 20 years before Eldredge's and Gould's first essay on the subject³.

The difference between Gould and Eldredge on the one hand, and (for example) G. G. Simpson on the other (see ref. 2), is that 'punctuated equilibria' came to be explained by Mayr's peripatric theory of speciation⁴, whereas Simpson's 'quantum jumps' were attributed to the invasion of new adaptive zones⁵. Gould and Eldredge agree, however, that speciation by itself is unlikely to promote morphological change. However, they do not mention the most obvious explanation of a historical association between speciation and change, namely that both processes are facilitated when organisms find themselves in new niches or habitats. Of course this explanation is inimical to Gould and Eldredge's belief that much of macro-evolution is due to species sorting. If both speciation and morphological change come about through the invasion of new habitats, the direction of macro-evolution will be largely determined by the standard processes of natural selection acting within populations. Eldredge and Gould's theory then becomes indistinguishable from Simpson's. Perhaps we should, more appropriately, be celebrating the fiftieth anniversary of G. G. Simpson's 'quantum jumps'.

Bryan Clarke

*Department of Genetics,
University of Nottingham,
Queens Medical Centre,
Nottingham NG7 2UK, UK*

SIR — I note that Gould and Eldredge in their self-congratulatory Review¹ on punctuated equilibrium cited my book⁶ as declaring punctuated equilibrium as being either dead or trivial. They got it wrong. I regard punctuated equilibrium as vacuous, since it was predicated on the straw man of phyletic gradualism, which, in turn, was used to set up an inappropriate dichotomy. The palaeontological literature has always presumed sudden change

to establish biostratigraphic relationships⁶.

Gould and Eldredge give credit to punctuated equilibrium for an understanding of evolutionary processes above the species level, but the concept is unnecessary for species selection or sorting to work⁷. They misleadingly quote Williams⁸ who acknowledges that evolutionary processes might work above the species level but does not accept punctuated equilibrium at all. Authors they cite as supportive in fact see no particular challenge of punctuated equilibrium to evolutionary theory⁹. Indeed, Gould and Eldredge have devolved their claims of punctuation from an "alternative" to being "complementary". Because the complement involves placing an artificial border in a complex array, nothing of substance has been really added, even if a couple of examples now exist, and Gould and Eldredge have apparently finally admitted it.

Jeffrey Levinton

Department of Ecology and Evolution,
State University of New York,
Stony Brook, New York 11794, USA

GOULD AND ELDREDGE REPLY — We accept the grudging support of our oldest and severest critics in gradual increments. How lovely that Levinton now grants us "a couple of examples", which must at least make punctuated equilibrium trivial and not entirely vacuous. We stated ourselves¹ that species selection and sorting do not logically require punctuated equilibrium, but it would be both contrary to the historical record (and ungenerous in the extreme) to deny that this important discussion about macroevolutionary theory was initiated and primarily nurtured in the context of 20 years' debate about punctuated equilibrium (a pattern that compels attention to species sorting as the primary mechanism of evolutionary trends).

We did not quote George Williams as a supporter of punctuated equilibrium, but neither does he share Levinton's sourness, as Williams wrote in the book we cited: "I will assume here that there is some factual validity to the pattern they [Gould and Eldredge] describe."⁷ We did not cite punctuated equilibrium as an alternative theory of evolution (as Levinton states), but as an alternative to the pattern of phyletic gradualism, once viewed as canonical and nearly exclusive. The title of our original 1972 article³ reads: "Punctuated equilibria: an alternative to phyletic gradualism". Alternatives are often complements, so what is Levinton's beef?

Clarke uses the venerable stratagem of "old hat" or "all been said before". But punctuated equilibrium is not the observation that evolution may occasionally move in spurts. Of course such an idea is old hat; Clarke needn't take us back to 1949 for this; why not Huxley's famous comment to Darwin in 1859: "you load yourself with

an unnecessary difficulty in adopting *natura non facit saltum* so unreservedly."

Punctuated equilibrium is a theory that attributes this pattern of spurt and stasis neither (1) to imperfections of the fossil record in a truly gradualistic world, nor (2) to such theories of occasional anagenetic rapidity as Simpson's important

1. Gould, S. J. & Eldredge, N. *Nature* **366**, 223–227 (1993).
2. Jepsen, G. L., Simpson, G. G. & Mayr, E. (eds) *Genetics, Paleontology and Evolution* (Princeton University Press, 1949; and Atheneum, New York, 1963).
3. Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T. J. M.) 82–115 (Freeman, San Francisco, 1972).
4. Mayr, E. *Animal Species and Evolution* (Harvard University Press, Cambridge, Massachusetts, 1963).
5. Simpson, G. G. *Tempo and Mode in Evolution* (Columbia University Press, New York, 1944).
6. Levinton, J. S. *Genetics, Paleontology, and Macroevolution* (Cambridge University Press, New York, 1988).
7. Vrba, E. S. *Science* **221**, 387–389 (1983).
8. Williams, G. C. *Natural Selection: Domains, Levels, and Challenges* (Oxford University Press, New York, 1992).
9. Fortey, R. A. *Sci. Prog. Oxf.* **72**, 1–19 (1988).

Longevity and testosterone

SIR — Nieschlag *et al.*¹ concluded that testosterone may not be responsible for the difference in lifespan now present between men and women. But their data do not appear to me to support this statement.

With the standard deviations reported in ref. 1 the statistical analysis would not have the power to detect a difference in lifespan even of 5 years due to the small sample number. But anyhow there is a more important objection. The 'castrati' singers were almost exclusively from Italy², as were many famous singers in that period (this was the age of the Bel Canto³). I therefore consulted an Italian life table. Because there are no accurate data for the seventeenth and the eighteenth century I examined a historical period in which the life expectancy was similar to that of the group under study for an individual who had reached the age of 30.

In the year 1891 the expected age of death of an Italian man who had reached age 30 was 64.7 years and that of a woman 65.0 years⁴. The life expectancy in 1891 was indeed almost equal between Italian men and women at all ages. Even if the criteria with which I have selected this population are open to argument, it is clear that we cannot assume that there was a difference in mortality between the sexes at the time of the castrati.

It is only since the First World War that increased female longevity is clearly present in all the developed countries⁵. There are many examples of small or inverted differences in the less developed countries even today⁶ which cannot simply be explained by a high incidence of maternal death⁵. The most important cause of the

hypothesis⁵ of quantum evolution, but to speciation as a process of branching, characteristically occurring at geologically instantaneous rates — with trends then explained not as anagenetic accumulation, but as differential success by species sorting.

It has been remarked that new theories tend to go through three stages: dismissed first as false, then rejected as contrary to "religion" (read accepted ideas), and finally embraced as true but long known after all. How nice to be at that third stage already, when punctuated equilibrium has just become old enough to raise a glass legally to its own good health.

Stephen Jay Gould

Museum of Comparative Zoology,
Harvard University,
Cambridge, Massachusetts 02138, USA
Niles Eldredge
Department of Invertebrates,
American Museum of Natural History,
New York, New York 10024, USA

difference, when present, is an increase in men of premature death from heart disease⁷. But in our 1891 Italian population, heart disease was listed as a cause of death only in 6% of the cases⁸.

In conclusion, we cannot yet absolve testosterone. This is important not only for understanding the smaller incidence of coronary heart disease among women, but also because androgen hormones are sometimes taken for non-medical reasons (for example, enhancement of athletic performance) and their potentially damaging effects should not be underestimated.

Giovanni Paternostro

Department of Biochemistry,
University of Oxford,
Oxford OX1 3QU, UK

1. Nieschlag, E., Nieschlag, S. & Behre, H. M. *Nature* **366**, 215 (1993).
2. Sadie, S. (ed.) *The New Grove Dictionary of Music and Musicians* (Macmillan, London, 1992).
3. Scholes, P. A. (ed.) *The Oxford Companion to Music* 10th edn (Oxford University Press, 1970).
4. Preston, S. H., Keyfitz, N. & Schoen, R. *Causes of Death: Life Tables for National Populations* (Seminar Press, New York, 1972).
5. Stolnitz, G. J. *Population Studies* **10**, No. 1, 17–42 (1956).
6. UN Statistical Yearbook 1990/91 38th issue (New York 1993).
7. Johansson, S. *Cardiovasc. Clin.* **19**, 3–16 (1989).
8. Istituto Centrale di Statistica *Sommario di Statistiche Storiche Italiane 1861–1955* (Roma, 1958).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Muddleheadedness exposed

Donald Kennedy

Higher Superstition: The Academic Left and Its Quarrels With Science. By Paul R. Gross and Norman Levitt. Johns Hopkins University Press: 1994. Pp. 314. \$25.95, £21.50.

ONCE, near the height of the 'sixties revolution', I attended a meeting at which several scientists had been brought together to hear the 'concerns' of a group of students and others committed not only to ending the Vietnam War (an engagement prosecuted, in their view, with quite a lot of help from Stanford University) but also to creating a more open, less elitist university. There, a well-known colleague from one of the humanities departments asserted his confident belief that "people's photosynthesis" was likely to have a higher quantum yield than the capitalist variety.

I supposed that such thinking had probably died with the revolution. Indeed, most scientists — even those who spend much of their time, as I do, in nominally 'interdisciplinary' work — have never looked deeply into the literature of the postmodern humanities. The closest view we are likely to have comes through an occasional skirmish over Derrida or another of the major prophets in *The New York Review of Books* — exchanges apt to be long on rhetoric and short on the careful expression of views. So we academic scientists have retained a set of comfortable assumptions about our colleagues, including in particular the assumption that they share generally our belief in what John Searle has called the "Western rationalistic tradition". For science and scientists, this means simply that reality exists independently of human representations, and that we can — in Searle's words — "find true sentences, ideally in the form of precise theories, that are true because they correspond, at least approximately to [that] independently existing reality".

Well, we are in for a shock; and Paul R. Gross and Norman Levitt deliver it in *Higher Superstition*. It is a convincing, and therefore frightening, treatise on the degree to which antiscientific thinking has taken hold in a domain the authors call the "academic left". In this category they include those humanists and social scientists who are primarily concerned with cultural studies and who belong, generally, to the 'postmodern' school. The taxonomy is necessarily vague and indeed presents some problems, but these do not seriously damage, let alone disable, the authors' central claim.

The case they make has rescued me from a more recent and troublesome problem than the quantum yield of people's

photosynthesis. For a dozen years as president of a leading US university — whose faculty members, thankfully, all escape Gross and Levitt's fire — I not only carried out but also enthusiastically supported several policies that were unattractive to the right. I encouraged increased access for minority students, deployed without embarrassment the term 'multiculture' (before, however, it acquired its present load of political freight) and defended a modest broadening of the canon of Stanford's required course in Western Culture. These efforts were rewarded with predictable scorn in the editorial pages of the *Wall Street Journal* and the spate of tell-all books about US universities by the late Allan Bloom, Dinesh d'Souza and their like. These assaults were more effective than they should have been, in part because the academic left kept passing ammunition to the right. The follies of extreme cultural relativism, deconstruction and so on made a sensible

middle ground almost impossible to defend. What thoughtful advocate of women's rights cannot be made to wince at the notion of a 'feminist algebra'? What promoter of intercultural understanding does not want to seek a hiding-place when an 'Afrocentric' scholar argues for widespread pre-Columbian encounters between African and meso-American peoples?

To those of us adrift in the Bermuda Triangle of old-fashioned liberalism, Gross and Levitt have tossed a lifebelt. *Higher Superstition* is a tough-minded, stinging assault on all manner of scientific foolishness; yet it does not come from the right. It is neither an anti-intellectual nor an anti-academic book, and it draws careful distinctions between what is controversial yet worthwhile and what is pure nonsense. That is because the authors have made a responsible effort to dissect without damaging vital tissue, and because their central strategy is to damn their targets out of their own mouths.

It works. The scientist who has avoided any exposure to postmodern thinking in the humanities will be surprised to read:

...falsifiability is often put forward as a criterion for distinguishing between the truly scientific and the pseudo-scientific. But such a yardstick is no more objectively adequate and no less mythical a criterion than appeals to, say, aesthetic complexity



FRIENDS in high places — scientists have only just begun to explore the biology of the tropical rainforest canopy. In *The High Frontier*, Mark W. Moffett provides an elegantly written account of the ecology of this aerial continent. The book is packed full with his stunning — and often unusual — photographs of canopy architecture, animals and plants, taken while accompanying tree-climbing researchers working at study sites worldwide. Harvard University Press, \$24.95, £17.95 (pbk); \$39.95, £31.95 (hbk).

have been in the history of cultural criticism. Falsifiability is a self-referential concept in science inasmuch as it appeals to those normative codes of science that favor objective authentication of evidence by a supposedly objective observer. In the same way "aesthetic complexity" only makes sense as a criterion of demarcation inasmuch as it refers to assumptions about the supposed objectivity of categories like the "aesthetic" referred by institutionally accredited judges of taste [p.89].

Surprise may turn to horror when one learns that this passage comes from a serious, scholarly journal and was written by a tenured professor at Princeton. (The ghosts of John von Neumann and other sources of that university's scientific distinction will be relieved to learn that he has since left.) But the strange belief that the essential elements of rationalism in science are as arbitrary and self-referential as those involved in judging art is by no means limited to a few of the postmodernists. Gross and Levitt demonstrate that this conviction is widespread, and has a firmer foothold than those in the scientific community may have suspected.

Theirs is a difficult mission, and it is not surprising that here and there the authors stumble. They can be accused of selecting some easy victims and avoiding some tougher ones. Jeremy Rifkin, for example — though he may be tolerated by some leftward academics — is hardly part of any academic left; yet he rates 15 pages of index entries, more than any other author. And on the other side, although a few superb women scientists have made arguments for a form of scientific relativism — Ruth Hubbard comes particularly to mind — they are avoided in favour of less well defended targets.

A more serious problem arises in areas where the distinctions between what is anti-scientific and what is legitimate scientific dispute are harder to make. In the sixth chapter, appropriately titled "The Gates of Eden", the authors take on ecofeminism and other forms of breathless Green enthusiasm. They do well at this, but there are more problematic points at which the knife slips. For example, the section on global environmental climate change errs in understanding and sometimes mis-characterizing the problem. Their thermostat analogy for the global greenhouse is wrong. Anthropogenic carbon dioxide is not changing a set point, because there isn't one; rather, it is influencing the overall heat budget. Ozone depletion, similarly, is about not just an Antarctic hole but also a thoroughly documented effect on the entire layer. And in their assessment of environmental over-regulation, the authors give 'cost-benefit' analysis a respectful bow without any reference to the deep difficulties involved in balancing economics on the one hand and health and environmental quali-

ty on the other. In this section, and here alone, the authors sound more like advocates for a political doctrine than careful critics.

These defects are, as I have said, not serious lesions in what is a well-constructed and valuable polemic. Because it is so good, it promises to be highly controversial. So I hope the authors have set the stage for a serious engagement with the issue in the academy — but about that I have some misgivings. It would be a terrible disappointment if *Higher Superstition* simply propelled us into a noisy struggle with some scientists on one side and some members of the academic left on the other. This is a conversation that needs many participants — including, most especially, those who have chosen to ignore or shrug off what has become an embarrassment to serious scholarship.

The first word in this admirable book is

'muddleheadedness'. We are at a critical juncture in the relationship between universities and the societies that support them. Public discontent about what we do is not restricted to the price of tuition, or the contribution of research to regional economies. Above all we are expected to engage in clear thinking — it is the quality on which our place in the world depends. If we are seen to tolerate or excuse muddleheadedness, we surely cannot complain if the world stops taking us seriously. The debate generated by *Higher Superstition* must therefore be not merely about how structuralists view science; it has also to be about quality of mind and how we judge it. □

Donald Kennedy is in the Department of Biological Sciences and the Institute for International Studies, Stanford University, Encina Hall 200, Stanford, California 94305-6055, USA.

The nature of knowledge

Maurice Crosland

Auguste Comte: An Intellectual Biography, Volume I. By Mary Pickering. Cambridge University Press: 1993. Pp. 776. £45, \$49.95.

THE philosopher Auguste Comte claimed that science was the only valid knowledge, a facile assumption for those educated exclusively in science. As the founder of positivism he seems to be permanently associated with multi-volume works. Not

1941) and the present substantial volume, covering the period from Comte's birth in 1798 to the completion of his *Cours*, promises to be the first of two.

The volume is a development of a Harvard University thesis and represents more than a decade of hard work including extensive research in the Paris archives. Most of the philosopher's prolific correspondence has survived, as has other manuscript evidence. Comte tried to work systematically and his latest biographer has carefully followed his tracks. Comte was an influential figure and there is therefore a large secondary literature. All of this has been combed exhaustively. Comte spoke of his life as a novel, and indeed some of his biography may seem stranger than fiction. The author has a special interest in Comte's political theory, the origins and development of which she traces in great detail. She also shows great sympathy for Comte's wife, who long endured his megalomania.

Mary Evans

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Founder of positivism: Auguste Comte (1798–1857)

only did his own *Cours de philosophie positive* run to six volumes (1830–1842) but analyses of his work seem to inspire the multi-volume approach. The standard work on the early life of the savant has been Henri Gouhier's *La jeunesse d'Auguste Comte* (3 vols., Paris, 1933–

We are taken through the struggles of Comte's personal life from his early days when he was expelled from the Ecole Polytechnique. His rejection of all authority and his need for independence presented constant problems, which made it difficult for him to earn a living except by giving private lessons. Despite an argument by the author to the contrary, Comte seems to have suppressed his emo-

tions as a young man in favour of an extreme rationality inherited from the Enlightenment. He had little sympathy for the imagination and, in his theory (for him, a "law") of the three stages, he claimed that knowledge had passed through a theological stage and a metaphysical stage before arriving finally at a scientific or "positive" stage, based on well attested facts. No hypothesis should be admitted that could not be verified by observation. Thus Comte warned physicists against theories invoking the aether. Some of his scientist followers refused to accept a theory of atoms or molecular structure. Even before Comte there was a tendency in French science never to go beyond laboratory experience. This tradition was intensified by the influence of Comte's philosophy. Comte spread his ideas mainly through his writings, but at the age of 28 he also began to give a series of lectures. The reader, who has been carefully following the development of his philosophy, is suddenly plunged from the extremes of rationality into a personal maelstrom of irrationality. Comte went mad and it was two years before he recovered.

Like the early Saint-Simon, Comte believed that scientists should play an important part in creating the new social order. Yet his model scientist was not the specialist but the generalist. One of his criticisms of the French Academy of Sciences was that it was composed of eminent specialists. Indeed, this was the basis of the high standards it set. The Academy was doubly important in Comte's life. First, membership was considered the peak of any scientist's career and Comte, sure of his own genius, made several unsuccessful attempts to be elected. Although a competent teacher of mathematics, he could never lay claim to research qualifications. It was only as a philosopher of science that he could claim distinction and this was not an area recognized by the Academy. But the Academy also had the power of nomination to positions in certain higher educational establishments, including Comte's old school, the Ecole Polytechnique. Comte tried repeatedly to be nominated for a chair of mathematics but he had to be content with the junior position of *répétiteur*. When in 1838 he was also appointed as admissions examiner to the Ecole Polytechnique he finally achieved a reasonable income. In his early career, Comte's ideas attracted a handful of eminent scientists, but he increasingly antagonized them with his bizarre ideas and even personal attacks when they failed to give their unqualified support. There was much paranoia in Comte's dealings with his contemporaries.

Greatly concerned with the problem of social stability after the violence of the French Revolution, Comte consorted with both liberals and counter-revolu-

tionaries. Several of his predecessors had tried to make politics a science. Comte himself tried to create a whole science of society, which he first called "social physics" but which from 1839 became "sociology". Unlike his predecessor Condorcet, however, also a mathematician, he did not approve of the quantification of social phenomena, which is strange, considering that positivism is today often associated with statistical analysis.

This book therefore provides an exhaustive account of Comte's life and work and is fully documented. But the inclusion of a chronological table, giving the most important events in his life, would have been useful. For those with a special interest in the history of ideas, this is an important book that should replace Gouhier's as the standard work on Comte's early life. □

Maurice Crosland is in the Unit for the History of Science, Physics Laboratory, University of Kent, Canterbury CT2 7NR, UK.

Only connect

Adam Kuper

Companion Encyclopedia of Anthropology: Humanity, Culture and Social Life. Edited by Tim Ingold. *Routledge: 1994. Pp. 1,127. £85, \$150.*

"MOST academic disciplines and their boundaries are, in fact, the fossilized shells of burnt-out theories," writes Tim Ingold in the introduction to this volume, "and in this, anthropology is no exception". That can scarcely be denied. Anthropology departments still often yoke together evolutionary biologists, archaeologists and cultural and social anthropologists, who commonly cohabit in a state of mutual mistrust and incomprehension. The original justification for this uneasy ménage is now generally regarded as a terrible error.

The pioneer anthropologists by and large agreed that race was the key to human variation. Race determined culture. Moreover, they were confident that the human races and their cultures could be ordered in an evolutionary series. Finally, just as contemporary apes represented the remote ancestors of the human species, so the ways of life of the 'lower races' represented the ancient past condition of 'civilized' populations.

These shared assumptions allowed the anthropologists to define a unified project, and to divide the field in a coherent way between emerging specialisms. Evolutionary biologists would trace the genealogy of the species and uncover the specific characters of each race; archaeologists would investigate the fossils of ancient cultures; ethnologists would

study the habits and customs of contemporary 'primitive' peoples.

This programme came unstuck in the first half of the twentieth century. It was established that race, culture and language were independent, and that racial classifications were extremely problematic and had no bearing on intelligence, moral dispositions or the capacity to master cultural skills. It turned out to be extremely difficult, if not impossible, to reconstruct the early social and cultural history of the species. There were no 'types' of culture that could be ranged in evolutionary 'stages'. Indeed, the most evident patterns of cultural variation were regional, and they were not necessarily associated with modes of production.

As these grand research programmes unravelled, the various anthropological specialisms developed their own agendas. Although sometimes still thrown together in the same university departments, many specialists felt that they had more in common with colleagues in other disciplines, even other faculties, than with their immediate neighbours. Nevertheless, there were several problems on which specialists could still usefully collaborate; and some still hankered after a single, unifying theory of human nature and cultural variation.

Ingold half promises that this encyclopaedia will provide a new synthesis. The subject of the book "is not anthropology but human life, and...its orientation is to the future rather than the past... While conceived as a work of reference, its aims go far beyond that: namely to lay the foundations for an integrated and synoptic perspective on the conditions of human life that is appropriate to the challenges of the next century." What in fact it provides is a fair review, by a variety of competent specialists, of some of the main subfields of anthropology. Ingold himself, in a series of introductory essays, attempts to make connections, some suggestive, others rather too rapidly sketched, but his contributors abundantly (though by no means exhaustively) illustrate the diversity of theoretical perspectives currently in play in the human sciences — or rather in the set of human sciences conventionally represented in old-fashioned anthropology departments. There is little on psychology or even on the brain, omissions that will surprise many scientists with rather different notions of what challenges the human sciences will face in the next century.

The book is arranged in three sections, corresponding to the traditional specialisms of physical anthropology, cultural anthropology and social anthropology. (Archaeology is notably under-represented, allowing Ingold to claim that technology has not been taken seriously enough by anthropologists.) Part I, "Humanity", has chapters on humans and

animals, human evolution, tools and early modes of production, diet, disease and demography. Part II, "Culture", is loosely ordered by the contested idea that cultures should be treated as systems of symbolic action, but the section concludes with an essay on ethnicity and nationalism that might have fitted more easily into Part III, "Social Life". This last section contains an even more eclectic set of essays, ranging from "Social aspects of language use" to "The nation state, colonial expansion and the contemporary world order".

Ingold concedes that a number of alternative topics might equally well have been treated, and there is not much point in listing some obvious gaps, although it is perhaps worth asking why — for there are fewer than 40 chapters in all — it was thought necessary to carry articles on "Perceptions of time", "Myth and metaphor", "Ritual and performance", "The anthropology of art" and "Music and dance". The choice of articles was bound to be idiosyncratic, however; despite the promise of the Introduction, the contributors do not develop a new synthesizing vision that might have dictated a more

coherent selection. Most of the contributors present their material within the conventional frameworks of their sub-disciplines. A few authors do advocate various synthesizing theories, but there is little common ground between, for example, F. J. Odling-Smee's sophisticated essay on niche construction and evolution and R. I. M. Dunbar's ideas about the broad differences between human and animal sociality; and neither is obviously relevant to what, for instance, Simon Roberts has to say about law and dispute processes, or James Weiner about myth and metaphor or Peter Worsley about colonialism.

This book is not an encyclopaedia of anthropology, much less of the human sciences, nor does it offer a new synthesis. It is a useful collection of generally authoritative reviews of some of the various projects in which anthropologists are currently engaged, occasionally (too occasionally) in a fruitful interdisciplinary fashion. □

Adam Kuper is in the Department of Anthropology, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

golden age of space exploration — also hinders the discussion of the origin and early history of Earth's atmosphere. Comins suggests that the 15 per cent difference between the amounts of carbon dioxide on Venus and Earth results from the Moon-forming impact — a rather small difference for such a large event. And how are we to account for the tiny amount of carbon dioxide on Mars?

In his treatment of the evolution of life on these alternative worlds, Comins becomes obsessed with circadian rhythms. He worries about the survival of life on planets with short rotation periods or seasonally lengthened days and nights, but he tells us nothing about the Eskimos, penguins, polar bears, seals and other organisms who live quite happily on our own planet in regions with diurnal rhythms very different from those at lower latitudes.

It is not easy to discern the book's intended audience. It would not be too difficult for an educated lay reader, preferably with a college degree. Such a person may be put off by phrases such as "It lays on the ecliptic" yet would probably know, unlike the author, that it is hydrogen sulphide and not sulphur dioxide that smells like rotten eggs.

Does any of this really matter? Don't we simply want to be entertained? After all, Van Gogh made marvellous paintings portraying the Moon in impossible phases. But this book sets out to be scientific, so it must be judged by different criteria. What if there were a second edition, thoroughly refereed and carefully edited? Now that might make our circadian rhythms sing. □

Tobias Owen is in the Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA.

Speculative astronomy

Tobias Owen

What if The Moon Didn't Exist? Voyages to Earths That Might Have Been. By Neil F. Comins. *HarperCollins: 1994. Pp. 315. \$20.*

WHAT if this book didn't exist? We would miss the author's playful speculations about alternative Earths, about the ways in which our planet might have evolved if circumstances surrounding its origin and early history had been different. But we would be spared a lot of misinformation.

Among the alternative worlds Comins considers are Earth with no Moon, Earth tipped on its side, Earth with a closer Moon and Earth with a more massive Sun. He includes a discussion of possible stellar catastrophes: a nearby supernova explosion, collision of Earth with a tiny black hole and a near-miss by a passing star. In the last chapter he addresses an imminent earthly concern: the depletion of the ozone layer.

The success of such a book depends on a combination of good writing, fresh ideas and a solid grasp of basic science. Comins is most successful when writing about stellar variations. But his discussions of planetary phenomena are often full of errors and his flights of fancy over the evolution of life on Earth seem unduly influenced by a few narrow viewpoints.

The past 30 years of planetary exploration have vastly increased our understanding of our nearest neighbours in space.

Unfortunately, little of this new knowledge penetrates the book. We find anomalies such as this: "Jupiter and Saturn . . . rotate so rapidly that the friction between their surfaces and atmospheres pulls the air into narrow streaming belts". It is well known that these giant planets do not have surfaces. The appearance Comins describes results from motions confined to the atmospheres. The discussion of the Roche limit is murky at best. Contrary to the author's assertion that the rings of Saturn "are all inside Saturn's Roche limit", the G and E rings are outside. Frozen methanes and ammonia are not major components of comet nuclei: they may not be present at all.

The title chapter on the development of Earth with no Moon makes no mention of Mars and Venus, two other planets with either negligible moons (Mars) or none (Venus). So in asserting that a moonless Earth would have a rotation period of 6 hours, the author fails to explain why the (retrograde!) rotation period of Venus is 243 days, while that of Mars is about the same as Earth's. These variations can be explained through the influence of impacts by large planetesimals. The author correctly identifies such an impact as the source of our Moon, but he ignores its effect on Earth's rotation rate and the inclination of our planet's rotational axis.

A lack of familiarity with the importance of impacts on planetary evolution — one of the great insights provided by the

New in paperback

Great Essays in Science by Martin Gardner (formerly published as *The Sacred Beetle*. . .). Prometheus, \$16.95, £14.50. Thirty-four of the best, including I. Asimov, C. Darwin, A. Einstein, S. J. Gould, L. Thomas, H. G. Wells and A. Huxley.

Apprentice to Genius: The Making of a Scientific Dynasty by Robert Kanigel, with a new epilogue. Johns Hopkins University Press, \$14.95, £12.50. An absorbing study, first published in 1986, of the mentor-protégé chain running through a dynasty of biomedical scientists working at the National Institutes of Health and the Johns Hopkins University.

Shadows of Forgotten Ancestors: A Search for Who We Are by Carl Sagan and Ann Druyan. Ballantine, \$14. An ambitious and wide-ranging popular account of the origins of life on Earth.

Effectiveness of bioremediation for the Exxon Valdez oil spill

James R. Bragg*, Roger C. Prince†, E. James Harner‡ & Ronald M. Atlas§

* Exxon Production Research Co., PO Box 2189, Houston, Texas 77252-2189, USA

† Exxon Research and Engineering Co., Annandale, New Jersey 08801, USA

‡ West Virginia University, Department of Statistics and Computer Science, Morgantown, West Virginia 26506, USA

§ University of Louisville, Department of Biology, Louisville, Kentucky 40292, USA

The effectiveness of bioremediation for oil spills has been difficult to establish on dynamic, heterogeneous marine shorelines. A new interpretative technique used following the 1989 Exxon Valdez spill in Alaska shows that fertilizer applications significantly increased rates of oil biodegradation. Biodegradation rates depended mainly on the concentration of nitrogen within the shoreline, the oil loading, and the extent to which natural biodegradation had already taken place. The results suggest ways to improve the effectiveness of bioremediation measures in the future.

MAJOR oil spills illustrate the need for effective and environmentally safe cleanup treatments. The *Exxon Valdez* spill of approximately 41 million litres of Alaskan North Slope crude oil in Prince William Sound, Alaska in March 1989 resulted in contamination with oil of ~2,000 km of rocky intertidal shorelines within the Sound and the Gulf of Alaska. Bioremediation was used extensively to accelerate the natural degradation of residual oil, thereby mitigating the ecological impact on intertidal communities. It involved the application of nitrogen-containing fertilizers to stimulate growth of indigenous hydrocarbon-degrading microorganisms within the intertidal zones; no microbial seeding was employed. Shorelines were treated with Inipol EAP 22 (7.4% N, 0.7% P)¹, an oleophilic liquid fertilizer, and Customblen (28% N, 3.5% P), a slow-release granulated fertilizer. Approximately 50,000 kg of nitrogen and 5,000 kg of phosphorus were applied to shorelines over the summers of 1989–92, making this the largest bioremediation project to date.

Pioneering, small-scale field trials^{2,9}, most completed before the *Exxon Valdez* spill, successfully demonstrated bioremediation potential, showing that addition of fertilizers could stimulate microbial growth and oil degradation. But it is difficult to quantify degradation for oil samples from the open marine environment, and most studies were unable to show statistically significant effects of fertilizer addition. In 1989, field tests on *Exxon Valdez* oil residues by the US Environmental Protection Agency (USEPA) provided striking visual evidence that bioremediation works¹⁰. Treated areas seemed significantly cleaner than adjacent untreated areas within a few weeks of fertilizer application. Laboratory testing^{11–13} confirmed that Inipol EAP 22 was not an effective agent for physically removing oil from sediments, and that oil removal was biological. Unfortunately, because sediments and oil loading were heterogeneous, extensive field testing based primarily on gravimetric techniques did not provide convincing evidence that bioremediation was effective^{14,15}. Therefore, a new interpretative approach was needed.

In 1990, a new field monitoring programme was implemented that included extensive chemical analyses of the oil; the results are shown here. We demonstrate that measuring changes over time in the oil composition relative to a stable, high-molecular-weight hydrocarbon present in the oil when spilled allows quantification of the rate and extent of the oil biodegradation with high levels of statistical confidence. For the *Exxon Valdez* oil, we found that hopane, a multi-ring saturated hydrocarbon, was not biodegraded and was a good conserved standard.

BOX 1 Statistical comparison of fertilized versus control

THE most straightforward way to interpret the effectiveness of bioremediation is to compare changes with time in the amounts of key hydrocarbon species in adjacent fertilized and control plots. We used the following first-order model for this comparison:

$$C_n(t) = (\alpha_1 + \phi_1 \chi) e^{(\gamma_1 + \phi_2 \chi)t} \varepsilon_1 \quad (1)$$

where $C_n(t)$ is the amount of a hydrocarbon component (for example, TGCDHC) relative to hopane at time t , χ is an indicator variable which is set to 1 when fertilizer is applied and to zero otherwise, α_1 , ϕ_1 , ϕ_2 and γ_1 are parameters determined from regression, and ε_1 is the assumed multiplicative error term. The rates of change for fertilized and control stations are:

$$\frac{d}{dt} C_n(t) = (\gamma_1 + \phi_2 \chi) C_n(t) \quad (2)$$

The initial composition, $C_n(0)$, is estimated more accurately from the model using the multiplicative term $(\alpha_1 + \phi_1 \chi)$ rather than from the limited data at day zero. The difference between the rate constants is estimated by ϕ_2 . The principal null hypothesis, that $\phi_2 = 0$, implies that the rate constants for fertilized and control treatments are equal.

For simplicity, regressions were performed on the linearized (logarithmic) form of equation (1). Figure 2 shows plots of the linearized models for TGCDHC/hopane, TRHC/hopane and TPAH/hopane for the subsurface of site KN-135. Only data up to day 72 for the control portion were included in the regression because fertilizer was applied to the control as well as the fertilized portion of KN-135 on that date²³. Over the first 70 days, there was no significant change in the oil composition in the untreated beach. The slopes of the control are at, or near, zero, suggesting that for this particular location and low level of background nitrogen, the rate of degradation was slow by the summer of 1990. For all three component ratios, the slopes of the fertilized regression lines were different from zero and had high significance, with values of R^2 (fraction of the total variation explained by the model) as shown in Fig. 2 and with P -values < 0.01 for all three. The test that fertilizer significantly enhanced bioremediation relative to the control beach, that is, $\phi_2 > 0$, was highly significant ($P < 0.01$) for all three hydrocarbon groups. Finally, the distribution of residuals showed that the model was appropriate for the data.

Monitoring hydrocarbon losses relative to hopane provided benchmark confirmation of oil biodegradation, and showed that adding fertilizers could accelerate the rate of oil removal by a factor of five or more. Perhaps as significant is the finding that the rate of oil biodegradation is a function of the nitrogen concentration maintained in the pore water of the intertidal sedi-

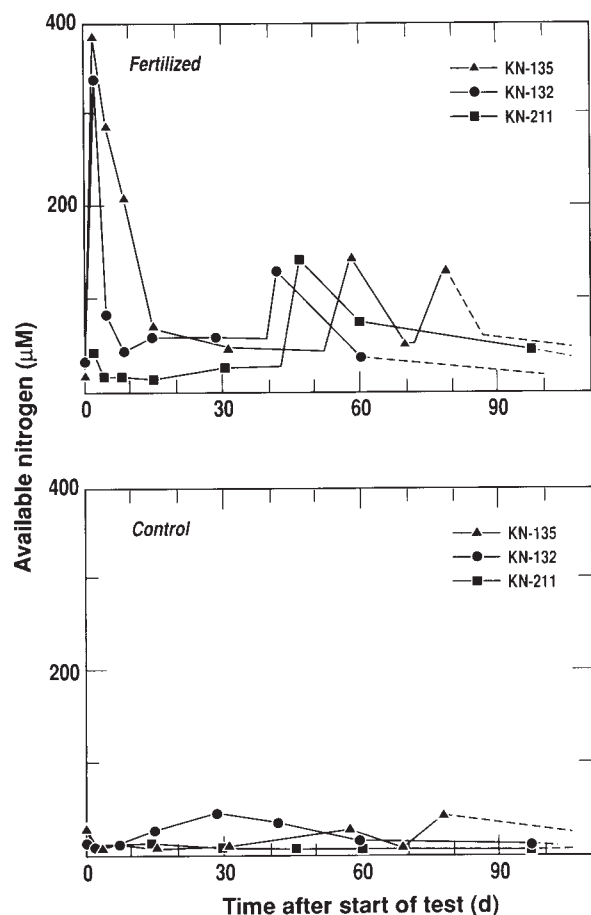


FIG. 1 Total available nitrogen concentration in interstitial pore water from fertilized (top) and control (bottom) sections of the three monitoring sites. Total nitrogen includes the sum of ammonium and nitrate, measured independently, plus the total organic nitrogen measured by a micro-Kjeldahl method minus the ammonium measurement^{16,18}. Although the fertilizers contain phosphate, no significant increase in this nutrient was detected¹⁶. Data points are means of three or six analyses; typical standard error for means was $\pm 15\%$ of mean or less. Fertilizer was applied on beaches at the following times: KN-135 on days 0, 53 and 72; KN-211 on days 0 and 44; KN-132 on days 0 and 40. For fertilized sites, nitrogen concentrations between days 30 and 60 reflect interpolated declines to day of second fertilizer application. Dashed lines indicate extrapolation beyond last measurements; results were found to be insensitive to the manner of extrapolation.

ments. This suggests that the effectiveness of bioremediation can be improved by making real-time measurements of nutrient levels in sediments to ensure that adequate, but safe, levels of nutrients are maintained during treatment.

Field data

In 1990, a field monitoring programme was initiated jointly by the USEPA, Alaska Department of Environmental Conservation and Exxon¹⁶ to provide statistically valid evidence of bioremediation effectiveness based on changes in oil composition as well as microbial and physical measures. Tests were conducted at three separate sites on Knight island in Prince William Sound. Site KN-135 was a low-energy beach that contained both surface

and subsurface oil, site KN-211 was a high-energy beach containing primarily subsurface oil, and site KN-132 was a low-energy beach with only surface oil. KN-132 and KN-135 were treated with Customblen and Inipol; KN-211 received only Customblen. All three sites were more heavily oiled than most shorelines treated in 1990 and, except for one small portion of KN-211, none had undergone bioremediation in 1989.

Each site was subdivided; half received fertilizer and half remained as an untreated control. Three sampling wells (pipe extending 70 cm into the subsurface) were placed in the fertilized portion of each site and three in the control. Samples were collected just before fertilizer application and periodically up to 109 d later. At low tide on the sampling dates, interstitial pore

TABLE 1 Initial concentration of oil on beaches and subsequent losses

Beach	Oil load* (mg oil per kg sediment)	Total cumulative hydrocarbon loss by weathering and biodegradation relative to North Slope crude oil (wt%)†	Natural biodegradation rate relative to oil 30% depleted of volatile compounds (% yr ⁻¹)‡	Natural biodegradation rate relative to oil 30% depleted of volatile compounds (g per kg sediment yr ⁻¹)§
KN-135 surface	7,190 $\pm$ 1,790	38 $\pm$ 2.4	10 $\pm$ 2.1	0.61 $\pm$ 0.22
KN-135 subsurface	6,653 $\pm$ 925	33 $\pm$ 0.8	3.8 $\pm$ 1.0	0.27 $\pm$ 0.10
KN-211 surface	1,990 $\pm$ 925	59 $\pm$ 4.7	36 $\pm$ 4.0	1.43 $\pm$ 0.62
KN-211 subsurface	13,900 $\pm$ 2,630	47 $\pm$ 2.5	21.0 $\pm$ 2.1	3.6 $\pm$ 0.80
KN-132 surface	4,730 $\pm$ 2,170	67 $\pm$ 4.3	45.8 $\pm$ 5.3	3.41 $\pm$ 1.09

Values given are arithmetic means ± 1 standard error.

* Whole oil (total extractable hydrocarbon) was solvent-extracted from the sediment sample and measured gravimetrically.

† Based on the difference between hopane concentration in whole Alaska North Slope oil spilled in March 1989 and hopane in whole-oil residue on the shoreline in May 1990.

‡ Assumes 30% of hydrocarbons were lost by physical weathering (vaporization/dissolution) before May 1989, and all subsequent losses until the start of the test in May 1990 were due to biodegradation; it is an estimate only because the total extent of weathering is unknown. Calculated from hopane concentrations in whole-oil residue in May 1990 and in Alaska North Slope oil with 30% of hydrocarbons removed.

§ Based on the estimated natural biodegradation rate (% yr⁻¹) (from adjacent column) and measured oil load.

water was withdrawn from each well, and three surface and three subsurface sediment samples were collected around each well¹⁶. Surface sediment samples were collected 2–5 cm below the pebbles; subsurface samples ~30–35 cm deeper. Fluid was analysed for fertilizer nutrients, dissolved oxygen, salinity, pH and temperature. Each sediment sample was cultured for microbes, and solvent extracted to measure oil loading. The three extracts for each well were then combined for a single compositional analysis by gas chromatography and mass spectroscopy (GC-MS)¹⁷, and oil loadings were averaged. At each sample date this yielded three values from the control and three from the fertilized portion for each sampling depth at each beach.

Fertilizer concentrations (Fig. 1) varied considerably over time as expected in a dynamic intertidal environment. Elevated concentrations of nitrogen were found in interstitial waters of beaches KN-135 and KN-132 after all fertilizer applications. At KN-211, the first application was unsuccessful, and nutrient measurements in offshore waters showed that much of the fertilizer was washed away only at this one site and only after the first application¹⁶. Subsequent fertilizer application there was successful (Fig. 1).

Hydrocarbon-degrading bacteria from fertilized portions of all three beaches showed statistically significant increases (up to 18-fold) in mineralization of radioactively labelled (¹⁴C) hexadecane and phenanthrene^{16,18}. Microbial cultures of all samples showed large numbers of bacteria (10^5 – 10^7 heterotrophs per g sediment), with ~10% being hydrocarbon degraders. Cultures generally showed higher numbers of bacteria at fertilized sites, but differences between fertilized and control sites were not always statistically significant¹⁸. Thus, microbiological data suggested that fertilizers stimulated growth of hydrocarbon degraders, but this did not provide truly convincing proof that the oil was being consumed.

Dissolved oxygen in pore water showed a transient decrease following fertilizer application^{12,16}, but remained unchanged at control sites, which is consistent with enhanced oil biodegradation. However, oxygen was not limiting because anaerobic conditions were not reached in these highly porous sediments. Adequate oxygen supply is required for hydrocarbon biodegradation^{19,20}, and in low-permeability, fine-grain sediments oxygen transport could become limiting²¹ even if fertilizer nutrient levels are adequate.

Changes in oil chemistry

In other spills, pristane and phytane have been used as non-degraded markers, but these were rapidly biodegraded^{12,22,23}. We therefore selected hopane as an internal conserved standard^{16,24}; hopanes have been used to monitor oil weathering after previous spills²⁵. 17 α (H),21 β (H)-hopane (a C₃₀ pentacyclic triterpane biomarker²⁶), the most abundant hopane in Alaska North Slope oil, is resistant to biodegradation on the timescale of years^{24,26}. Further, hopanes have low solubility in water and low volatility, so they remain in the residual oil on shoreline sediments as other hydrocarbons are lost by evaporation, dissolution and biodegradation. Their concentration in the weathered/degraded oil relative to the unaltered oil provides a quantitative measure of mass loss of other hydrocarbons. The maximum evaporation of Alaskan North Slope oil was ~30% (ref. 27), and most evaporation occurred within two weeks of the spill. By 1990, changes in oil composition by evaporation or dissolution were negligible, as confirmed from data monitored on unfertilized beaches (presented later in this Article).

Mean oil loadings and estimates of the extent of hydrocarbon losses by weathering and biodegradation before fertilizer was added are shown in Table 1. Substantial heterogeneity in the amount of oil on individual sediment particles resulted in large standard errors for gravimetric oil loading (~19–46% of mean). Estimates of loss calculated from changes in hopane concentrations relative to fresh North Slope crude show much lower standard errors (2–8% of mean). Hence, the composition of oil in

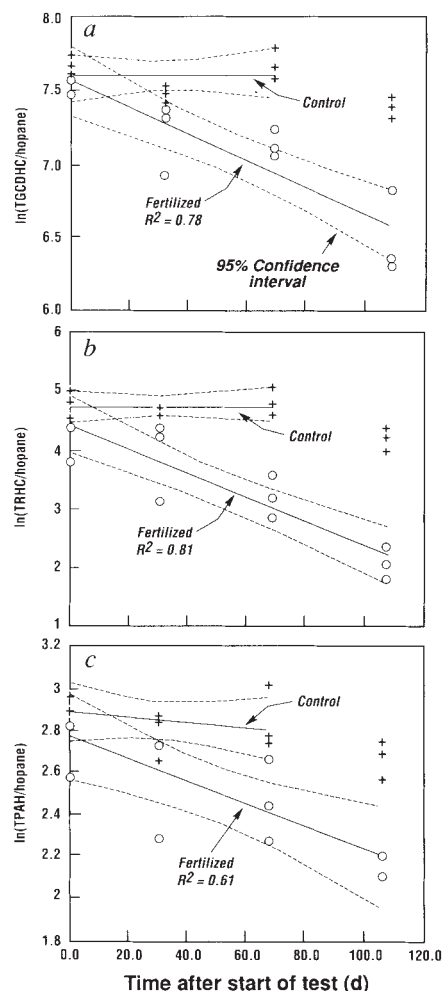


FIG. 2 Changes in the composition of a, total GC-detectable hydrocarbons (TGCDHC), measured with flame-ionization detection¹⁷, b, total GC-resolvable hydrocarbons (TRHC) measured with the same technique, and c, total polynuclear aromatic hydrocarbons (substituted and unsubstituted naphthalenes, phenanthrenes, dibenzothiophenes, fluorenes and chrysenes, TPAH) measured by mass spectrometry¹⁷, all with respect to hopane, in fertilized and control subsurface sediments on KN-135. GC data were corrected for extraction recovery efficiency and analytical sensitivities using surrogate tracers added to the sediment samples before extraction or analysis¹⁷; extraction efficiency was typically >99%; most analytical surrogate recoveries were in the range ~70–100%. Data were normalized using ratios of hydrocarbon mass/hopane mass for a fresh North Slope crude oil standard analysed at the same time. Data from the controls at day 109 are shown, but not fitted, because fertilizer was added to this part of the beach on day 72 (see Fig. 1).

replicate shoreline samples on a given date was less variable than the distribution of oil load on sediments. By assuming that oil mass lost by physical weathering (vaporization and dissolution) was 30% before May 1989, and negligible thereafter, we were able to estimate rates of natural biodegradation from May 1989 up to fertilizer addition in May 1990. (Unlike rates determined during the monitoring period when controls were available to confirm that losses from weathering were negligible, these background rates were only estimates because the total extent of weathering was unknown.) The rates varied considerably among locations; higher losses at KN-132 before adding fertilizer in 1990 may be due to locally higher background levels of nitrogenous nutrients (see Fig. 1, control).

The oil contains thousands of individual chemical species

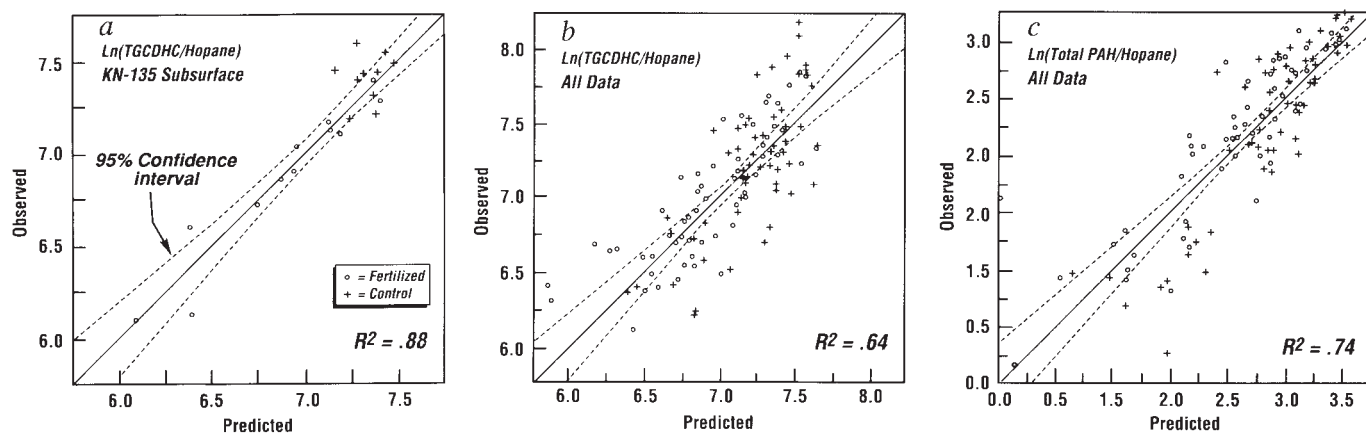


FIG. 3 Multiple regression fit of a, total GC-detectable hydrocarbons for KN-135 subsurface using equation (5) in box 2 (parameters were $\alpha_3 = 7.63$, $\nu_3 = 0.77$, $\delta_3 = -7.67$ and $\omega_3 = -0.00202$); b, total GC-detectable hydrocarbons for all three beaches, surface and subsurface, based on equation (3) in box 2 (parameters were $\alpha_2 = 6.21$, $\nu_2 = 0.862$, $\delta_2 = -0.000043$ and $\beta_2 = 0.159$); and c, total polynuclear aromatic hydrocarbons (substituted with 1–4 carbons and unsubstituted naphthalenes,

phenanthrenes, dibenzothiophenes, fluorenes and chrysenes^{17,23}) for all three beaches, surface and subsurface, based on equation (3) (parameters were $\alpha_2 = 0.479$, $\nu_2 = 1.47$, $\delta_2 = -0.0000015$ and $\beta_2 = 0.283$). If day zero data points for KN-211 and KN-132, suspected of having improper filtration of micrometre-sized mineral fines from oil samples before analysis¹⁶, are excluded the value of R^2 increases in b to 0.70 and in c to 0.77.

which degrade at different rates²⁸. To measure losses of important groups of hydrocarbons, we analysed changes over time in the ratio of their mass to the mass of hopane, including; total GC-detectable hydrocarbons (TGCDHC/hopane); total resolvable (nC_{10} – nC_{32} plus pristane and phytane resolved as specific peaks on the gas chromatograms) hydrocarbons (TRHC/hopane); and total polynuclear aromatic hydrocarbons (TPAH/hopane). The latter included substituted and unsubstituted naphthalenes, penanthrenes, dibenzothiophenes, fluorenes and chrysenes. Component compositions were determined by GC-MS¹⁷.

Compositions at the start of the test varied among locations and depths, but in a typical subsurface sample from KN-135, TGCDHC represented ~50–60% of the total oil mass, TRHC ~3%, polars (see below) ~20–30%, and the remainder was undetected non-eluting compounds. The oil sample was passed through an alumina column before GC-MS; the fraction retained was defined as polars. Analyses showed this fraction contained a complex mixture of asphaltenes and hydrocarbons

containing nitrogen, sulphur and oxygen. Polars in the oil increased as the oil became more biodegraded²³, suggesting that the polars fraction could serve as a measure of the extent of biodegradation. Earlier studies^{28–30} showed that polar hydrocarbons are more resistant to biodegradation than non-polars, and our laboratory data^{11,12} showed that as the polar content of biodegraded North Slope oil approached 60–70% of the total mass, biodegradation slowed substantially.

Interpreting effectiveness

Changes over time in oil chemistry relative to hopane provided the key to determining statistically significant effects of fertilizer addition. The simplest test of effectiveness was a comparison of the rate of hydrocarbon consumption in adjacent fertilized and control sites (Box 1). The results for three hydrocarbon groups, TGCDHC, TRHC and TPAH, are shown in Fig. 2 for subsurface oil in beach KN-135. Equation 1, with the slopes and intercepts from the regression fits^{12,23}, show that over 109 d in the fertilized site ~63% of TGCDHC, 90% of TRHC and ~44% of

TABLE 2 Mean values of key factors affecting oil biodegradation

Variables*	KN-135 (subsurface)	KN-211 (subsurface)	KN-132 (surface)	KN-135 (surface)	KN-211 (surface)
Oil load (g per kg sediment)					
Fertilized	3.22 ± 0.61	11.60 ± 2.05	5.17 ± 1.04	6.76 ± 0.96	2.06 ± 0.68
Control	6.13 ± 1.52	16.23 ± 0.88	5.64 ± 1.49	9.78 ± 0.96	4.07 ± 0.67
Polars (%)					
Fertilized	30.0 ± 1.7	20.4 ± 1.7	47.0 ± 4.9	38.2 ± 2.2	43.5 ± 4.7
Control	21.8 ± 1.2	15.8 ± 0.8	51.0 ± 3.1	30.9 ± 2.2	31.1 ± 2.6
Average $\bar{N}$, cumulative average N concentration (μM)†	109	39.2	72.0	109	39.2
Fertilized					
Control	9.4	17.1	23.4	9.4	17.1
($\bar{N}$ /oil load) × 1,000 (μM per mg oil per kg sediment)†					
Fertilized	46.3 ± 11.9	6.9 ± 2.4	17.0 ± 3.3	13.5 ± 2.1	52.5 ± 12.9
Control	4.6 ± 1.1	1.0 ± 0.1	11.9 ± 4.6	2.0 ± 0.4	6.4 ± 1.7

* Values given are arithmetic means ± 1 standard error for all data over the three-month test.

† Standard error for measured nitrogen concentrations was ± 15%. Cumulative average nitrogen concentration, $\bar{N}$, was calculated for each sample by integrating the nitrogen concentration in Fig. 1 to date of sampling and dividing by cumulative time in days. Uncertainty introduced from interpolated values was not assessed.

BOX 2 Multiple-regression models

MULTIPLE-regression models add power for interpreting effectiveness, allowing data from one or more test sites, subsurface and surface, to be combined. One such model found to give good fit to the data was:

$$C_n(t) = \alpha_2 [1 - p(t)]^{v_2} l(t)^{\beta_2} e^{\delta_2 N(t)} \varepsilon_2 \quad (3)$$

where $p(t)$ is the polar fraction, $l(t)$ is the oil load and $N(t)$ is the total cumulative nitrogen, all measured at time t . Parameters α_2 , v_2 , β_2 and δ_2 are coefficients determined by least-squares regression, and ε_2 is the assumed multiplicative error term. The rate model for equation (3) is then:

$$\frac{d}{dt} C_n(t) = \left[\delta_2 N'(t) - v_2 \frac{p'(t)}{1 - p(t)} + \beta_2 \frac{l'(t)}{l(t)} \right] C_n(t) \quad (4)$$

where ' denotes the derivative with respect to time. The rate coefficient is a linear function of the average cumulative nitrogen, $\bar{N}(t)$, with an intercept which is an additive function of the changes in polars and oil load relative to their absolute amounts at time t . For fertilized shorelines, contributions from the terms containing polars and oil load are small compared to the $N'(t)$ term, so increase in rate from fertilizer addition is approximately proportional to the increase in $\bar{N}(t)$ (for example, estimates of $p'/(1-p) \approx 0.00167 \text{ d}^{-1}$ and $l'/l \approx 0.0065 \text{ d}^{-1}$ for fertilized KN-135 (ref. 23), combined with parameters for TGCDHC/hopane in Fig. 3 and data for $\bar{N}$ in Table 2 illustrate this point).

This model predicts biodegradation reasonably well for all three beaches, with positive values of v_2 and β_2 , and negative values for δ_2 ; that is, the ratio of hydrocarbon component/hopane tended to be larger for higher oil loads and oil with less polars, but smaller where there had been more total nitrogen. This is consistent with

chemical, physical and biological factors governing oil biodegradation.

An alternative model, which improves the fit of subsurface data for individual beaches but otherwise gives levels of confidence comparable to those from equation (3), is based on the amount of available nitrogen per unit of oil, that is, on the nitrogen-to-carbon ratio. In analytical form, this model is:

$$C_n(t) = \alpha_3 [1 - p(t)]^{v_3} e^{\delta_3 r(t) + \omega_3 t} \varepsilon_3 \quad (5)$$

where $r(t) = \bar{N}(t)/l(t)$ is the average nitrogen to oil-load ratio as function of time, t ; $p(t)$ is the polar fraction, α_3 , v_3 , δ_3 and ω_3 are coefficients determined by least-squares regression; and ε_3 is the assumed multiplicative error term. An advantage of this model is that time is an explicit variable, whereas in the model of equation (3) time was implicit in the total nitrogen available to time t . The corresponding rate equation is:

$$\frac{d}{dt} C_n(t) = \left\{ \omega_3 + \delta_3 r'(t) - v_3 \frac{p'(t)}{1 - p(t)} \right\} C_n(t) \quad (6)$$

Regression parameters and confidence levels for this model have been reported^{12,23} for several combinations of shoreline data and hydrocarbon groups, including loss of individual PAH groups (naphthalenes, phenanthrenes, dibenzothiophenes and chrysenes). Results show that the hydrocarbon component/hopane ratio is larger for oil with less polars, but becomes smaller as average nitrogen concentration per unit of oil load and time increase. These results again are consistent with the chemical and physical processes controlling oil biodegradation on shorelines.

the TPAH were consumed. The control showed no significant change until fertilizer was added at day 72; then losses were similar to those observed initially in the fertilized portion.

Biodegradation was extensive at KN-135, less was noted at KN-211, and no significant biodegradation was noted at KN-132. The simple model in Box 1 that considered only the presence or absence of fertilizer was not adequate for beaches other than KN-135, so other factors had to be considered. Most significant are the amount of nitrogen available to microbes, time, the amount of oil and its chemical composition (to account for previous weathering and degradation). Table 2 summarizes mean values measured for key parameters over three months of monitoring. Polars were higher in surface oil, perhaps a result of enhanced biodegradation from greater exposure to natural nutrients or oxygen, or chemical weathering³¹.

Box 2 summarizes two multiple-regression models used to interpret how these parameters affected bioremediation. The models allowed testing individual or combined data sets. Both give reasonable fits to all data, and they differ primarily in the way that nitrogen concentration is related to time. The oil load and nitrogen appear in different terms in the two models, but the effect of their interaction is the same—higher oil loads require more nitrogen for effective biodegradation—an intuitive result that appears consistent with the system chemistry and biology.

Figure 3 illustrates observed versus predicted hydrocarbon/hopane ratios. Figure 3a shows the TGCDHC/hopane ratio for the subsurface of KN-135. A value of 0.88 for R^2 is greater than the value of 0.78 for a fit of the same data by the simpler model in Box 1. Thus, accounting for variations in oil load and nitrogen concentrations among samples improves fit. Figure 3b and c shows fits of TGCDHC and TPAH by the model in equation (3) for all data combined from KN-135, KN-211 and KN-132, surface and subsurface. These results led us to conclude that the most important parameter controlling effectiveness was the cumulative amount of nitrogen available per unit of oil, but stimulation of biodegradation was most pronounced where the oil was relatively unweathered.

How effective was fertilizer addition? A simple ratio of slopes from regression lines for fertilized and control would give a reasonable estimate of the increase in biodegradation rate. But the slope of KN-135 control in Fig. 2 is near zero, and this would suggest an almost infinite increase at this test site. A better estimate can be obtained from the multiple-regression equations described in Box 2, which incorporate control data with high levels of statistical confidence.

Equation (4) in Box 2, with parameter estimates in Fig. 3b for TGCDHC/hopane and values of $\bar{N}$ (the cumulative average nitrogen concentration in the interstitial water) in Table 2, predicts that over 109 d the degradation rate of TGCDHC in KN-135 (fertilized) was about $0.45\% \text{ d}^{-1}$, compared to $0.052\% \text{ d}^{-1}$ in KN-135 control (before fertilizer was added at day 72), giving a rate enhancement of 8.6 times.

The model of equation (6) in Box 2 relates the rate of biodegradation to the average concentration of nitrogen per unit of oil and polars. Over short time spans, the change in concentration of polars within the oil is small and compared to terms containing the nitrogen/oil load, the derivative of polars with time in equation (6) is negligible. Therefore, the instantaneous rate of degradation is approximately proportional to the average amount of nitrogen per unit of oil. We can therefore use equation (6), parameters in Fig. 3a and the value of $\bar{N}$ /oil load from Table 2 to estimate the degradation rate resulting from different nitrogen/oil loads. For KN-135 subsurface it predicts that fertilizer increased biodegradation of TGCDHC by a factor of ~ 5.5 over control (for equal oil loadings).

This model helps to explain why bioremediation appeared effective in only some of the tests conducted by the USEPA^{14,15}. Although hopane was not measured by USEPA, we have used available nutrient data from tests on Disk island¹⁴ and Elrington island¹⁴ to estimate the ratio of nitrogen to oil load. Figure 4 shows estimated stimulation ratios from equation (6). Despite the high fertilizer application rate (in g m^{-2} of beach area) on the Disk island sites, low nutrient levels in the interstitial water suggest that not enough was retained to be effective²³. Plot sizes were small (9 m^2) compared to tests such as KN-135 ($2,000 \text{ m}^2$),

and dispersion and dilution may have dissipated nutrients. Although there may have been additional reasons for the lack of demonstrated effectiveness at Disk island, Fig. 4 suggests that little stimulation of degradation would be achieved by measured nutrient levels. In contrast, nitrogen levels from the sprinkler system on Elrington island were much higher²³; Fig. 4 indicates that this should substantially enhance biodegradation, as observed¹⁴.

Our analyses provide guidelines for bioremediation of other spills. Seeding with oil-degrading microorganisms was not needed in Alaska, where shorelines abounded with indigenous hydrocarbon degraders. Whether seeding may be needed in any marine shoreline remains to be determined. Regardless of the source of microorganisms, it is essential that adequate nutrients be provided. Our results suggest that optimum rates of degradation would be achieved by maintaining high, but non-toxic, levels of nutrients. Fertilizer application would probably be appropriate soon after oil reaches a shoreline.

Conclusions

Because shorelines and oiling levels are naturally heterogeneous, the optimum amount of fertilizer will vary with location, and the best dosage cannot be predicted *a priori*. Therefore, it is important to base application strategy on monitored nutrients rather than predetermined dosage. On-site monitoring of nutrients in sediment pore waters would provide practical, real-time guidance on amounts and frequency of fertilizer applications. This would help to ensure that nutrient levels are adequate, but safe for marine biota. The monitoring programme conducted in Alaska jointly by the USEPA, ADEC and Exxon incorporated comprehensive environmental monitoring, including tests for toxicity and potential stimulation of photosynthetic plankton¹⁶. No adverse effects were found, and it seems likely that more frequent fertilizer application would have been equally safe and even more effective. But although Fig. 4 suggests an unbounded upper limit on increased rate of biodegradation from added nutrients (because of limits of the data and models used), upper limits will be constrained by the maximum safe concentrations of nutrients.

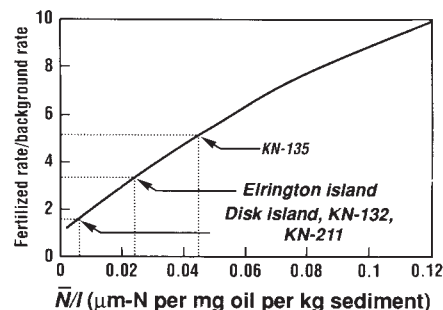


FIG. 4 The approximate increase in rate of biodegradation of total GC-detectable hydrocarbons relative to natural background (control) as function of the ratio of average nitrogen concentration to oil load ($\bar{N}/I$). Predictions are based on coefficient values for equation (6) shown in Fig. 3a. Estimated $\bar{N}/I$ are included for 1900 monitoring tests¹⁶ and for tests conducted by USEPA¹⁴.

There are limitations to this approach. Bioremediation is not likely to be effective on extensively degraded oil. We expect that once the polars content of the oil residue reaches ~60–70% of the total mass, nutrient availability will no longer be the limiting factor. Further, adequate oxygen must be available; bioremediation within anaerobic fine sediments such as mud flats or marshes would probably be effective only at shallow depths, but oil penetration in such sediments is also restricted. In Alaska, residue oil from the spill interacted with micrometre-sized mineral particles to form flocculated oil-in-water emulsions³² with large oil-water interfacial areas. In the absence of such emulsion formation, which microscopic studies showed enhanced access of microbes to the oil, the same rate enhancement from fertilizers may not always be realized.

Based upon the statistical analyses reported here, along with a substantial body of laboratory testing^{11–13} and field observations^{10,12,14}, bioremediation is an important treatment for oil spills on rocky intertidal shorelines of the type found in Alaska. □

Received 23 November 1993; accepted 21 February 1994.

1. Tellier, J., Sirvins, A., Gautier, J.-C. & Tramier, B. US Patent No. 4460692 (1984).
2. Sendstad, E. et al. in *Proc. 5th Arctic Marine Oilspill Program Tech. Semin.* Edmonton, Alberta, 331–340 (Environmental Protection Service, Ottawa, Ontario, 1982).
3. Sendstad, E., Sveum, P., Endal, L. J., Brattbakk, Y. & Ronning, O. L. in *Proc. 7th Arctic Marine Oilspill Program Tech. Semin.* Edmonton, Alberta, 60–74 (Environmental Protection Service, Ottawa, Ontario, 1984).
4. Halmo, G. in *Proc. 1985 Oil Spill Conf.* 531–537 (Am. Petrol. Inst., Washington DC, 1985).
5. Sveum, P. & Ladousse, A. in *Proc. 1989 Oil Spill Conf.* 439–446 (Am. Petrol. Inst., Washington DC, 1989).
6. Lee, K. & Levy, E. M. in *Proc. 1989 Oil Spill Conf.* 479–486 (Am. Petrol. Inst., Washington DC, 1989).
7. Lee, K. & Levy, E. M. in *Proc. 1991 Oil Spill Conf.* 541–547 (Am. Petrol. Inst., Washington DC, 1991).
8. Lee, K., Tremblay, G. H. & Levy, E. M. in *Proc. 1993 Oil Spill Conf.* 449–454 (Am. Petrol. Inst., Washington DC, 1991).
9. Rosenberg, E. et al. *Biodegradation* **3**, 337–350 (1992).
10. Pritchard, P. H. & Costa, C. F. *Envir. Sci. Technol.* **25**, 372–379 (1991).
11. Chianelli, R. R. et al. in *Proc. 1991 int. Oil Spill Conf.* 549–558 (Am. Petrol. Inst., Washington DC, 1991).
12. Bragg, J. R., Prince, R. C., Wilkinson, J. B. & Atlas, R. M. *Bioremediation for Shoreline Cleanup Following the 1989 Alaskan Oil Spill* (Exxon Co. USA, Houston, 1992).
13. Prince, R. C. et al. *Prepr. Div. Petrol. Chem. Am. chem. Soc.* **38**, 240–244 (1993).
14. Pritchard, P. H., Costa, C. F. & Suit, L. *Alaska Oil Spill Bioremediation Project* (Report No. EPLA/600/9.91/0.46a, b, US Envir. Protection Ag., Gulf Breeze, Florida, 1991).
15. Stone, R. *Science* **257**, 320 (1992).

16. Prince, R. C., Clark, J. R. & Lindstrom, J. E. *Bioremediation Monitoring Program* (Rep. to US Coast Guard) (Exxon Co. USA, US EPA, Alaska Dept of Environmental Conservation, Anchorage, 1990).
17. Douglas, G. S. et al. in *Contaminated Soils* (eds Kostecki, P. T. & Calabrese, E. J.) 1–21 (Lewis, Boca Raton, 1992).
18. Lindstrom, J. E. et al. *Appl. envir. Microbiol.* **57**, 2514–2522 (1992).
19. Hughes, D. E. & McKenzie, P. *Proc. R. Soc. B* **189**, 375–390 (1975).
20. Zobel, C. E. in *API/FWPCA Joint Conf. on Prevention and Control of Oil Spills* 317–326 (Am. Petrol. Inst., Washington, DC, 1969).
21. Gibbs, C. F. & Davis, S. J. *Microbial Ecol.* **3**, 55–64 (1976).
22. Pritchard, P. H., Mueller, J. G., Rogers, J. C., Kremer, F. V. & Glaser, J. A. *Biodegradation* **3**, 315–335 (1992).
23. Bragg, J. R., Prince, R. C., Harner, E. J. & Atlas, R. M. in *Proc. 1993 Int. Oil Spill Conf.* 435–447 (Am. Petrol. Inst., Washington DC, 1993).
24. Butler, E. L. et al. in *On-Site Bioreclamation* (eds Hinchee, R. E. & Olfenbuttel, R. F.) 515–521 (Butterworth-Heinemann, Boston, 1991).
25. Gundlach, E. R. et al. *Science* **221**, 122–129 (1981).
26. Peters, K. E. & Moldovan, J. M. *The Biomarker Guide* (Prentice-Hall, Englewood Cliffs, New Jersey, 1993).
27. Payne, J. R., Clayton, J. R., McNabb, G. D. & Kirstein, B. E. in *Proc. 1991 Int. Oil Spill Conf.* 641–654 (Am. Petrol. Inst. Washington DC, 1991).
28. Prince, R. C. *Crit. Rev. Microbiol.* **19**, 217–242 (1993).
29. Walker, J. D., Petrakis, L. & Colwell, R. R. *Can. J. Microbiol.* **22**, 598–602 (1976).
30. Walker, J. D., Colwell, R. R. & Petrakis, L. *Can. J. Microbiol.* **22**, 1209–1213 (1976).
31. Payne, J. R. & Phillips, C. R. *Envir. Sci. Technol.* **19**, 569–579 (1985).
32. Bragg, J. R. & Yang, S. H. in *Exxon Valdez Oil Spill: Fate and Effects in Alaskan Waters* (eds Wells, P. G., Butler, J. N. & Hughes, J. S.) (ASTM STP 1219, Am. Soc. Testing Materials, Philadelphia, in the press).

Physiological consequences of loss of plasminogen activator gene function in mice

Peter Carmeliet^{*†‡}, Luc Schoonjans[†], Lena Kleckens^{†‡}, Beverly Ream^{*}, Jay Degen[§], Roderick Bronson^{||}, Rita De Vos[¶], Joost J. van den Oord[¶], Désiré Collen^{†#} & Richard C. Mulligan^{*#}

* Whitehead Institute of Biomedical Research and the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

† Center for Molecular and Vascular Biology, University of Leuven, B-3000 Leuven, Belgium

§ Children's Hospital Research Foundation, University of Cincinnati, Ohio 45229-3039, USA

|| School of Veterinary Medicine, Tufts University, Boston, Massachusetts 01536, USA

¶ Laboratory of Histo- and Cytochemistry, University of Leuven, B-3000 Leuven, Belgium

Indirect evidence suggests a crucial role for the fibrinolytic system and its physiological triggers, tissue-type (t-PA) and urokinase-type (u-PA) plasminogen activator, in many proteolytic processes. Inactivation of the t-PA gene impairs clot lysis and inactivation of the u-PA gene results in occasional fibrin deposition. Mice with combined t-PA and u-PA deficiency suffer extensive spontaneous fibrin deposition, with its associated effects on growth, fertility and survival.

MAMMALIAN blood contains an enzymatic system, called the fibrinolytic system or the plasminogen/plasmin system, which has been claimed to play a role in a variety of phenomena associated with proteolysis, including blood clot dissolution (thrombolysis), ovulation, embryo implantation, embryogenesis, cell invasion and brain function¹⁻³. The fibrinolytic system comprises an inactive proenzyme, plasminogen, which is activated to the proteolytic enzyme plasmin by two physiological plasminogen activators, tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA). These are products of two separate genes with different genomic organization and chromosome localization².

t-PA is believed to be primarily responsible for removal of fibrin from the vascular tree²; it has a specific affinity for fibrin and produces clot-restricted plasminogen activation. The role of u-PA in thrombolysis is less well defined; it lacks affinity for fibrin and requires conversion from an inactive single-chain precursor to a catalytically active two-chain derivative². Whether other plasminogen activation pathways, such as the 'intrinsic' pathway (involving blood coagulation factor XII, high-molecular-mass kininogen, prekallikrein, and possibly u-PA⁴) contribute to clot lysis, remains to be established.

Urokinase-type plasminogen activator binds to a specific cell-surface receptor⁵ and is believed to be primarily involved in cell-mediated proteolysis by activation of latent matrix-degrading proteinases or growth factors^{6,7}. Its localization on the migrating front of cells suggests involvement in cell invasion through extracellular matrices^{3,6}. Directional proteolysis could play a role in physiological processes such as macrophage invasion⁸⁻¹⁰, ovulation^{11,12}, angiogenesis^{7,13} and wound healing^{7,14}, and in pathological cell invasion in neoplasia and metastasis¹⁵.

Fibrinolysis has been claimed to be involved in ovulation, fertilization, embryo implantation and embryogenesis, on the basis of its expression in ovulating oocytes^{8,10}, in migrating sperm cells¹⁶, in invading trophoblasts¹⁷ and in several primitive organs during embryogenesis^{18,19}. In addition, serine-proteinase

inhibitors and/or plasminogen-activator-specific antisera suppress ovulation¹² and embryo implantation²⁰.

The fibrinolytic system may also contribute to several pathological processes such as thrombosis²¹, atherosclerosis²², glomerulonephritis²³, acute respiratory distress syndrome²⁴, haemangioma formation¹³ and metastasis¹⁵. To date, genetic deficiencies of t-PA or u-PA have not been reported in man. Consequently, the implied role of the fibrinolytic system *in vivo* is deduced from correlations between fibrinolytic activity and (patho)physiological phenomena, which does not allow a causative role of the fibrinolytic system in these disorders to be established.

To investigate the role of t-PA- and u-PA-mediated plasminogen activation in development, reproduction, thrombolysis, thrombosis and macrophage function, we have generated mice with single and combined inactivation of the t-PA and/or u-PA genes. Mice with single deficiencies of t-PA or u-PA develop normally, are fertile and have a normal life span. t-PA-deficient mice had a reduced thrombolytic potential and increased incidence of endotoxin-induced thrombosis. u-PA-deficient mice occasionally developed spontaneous fibrin deposits in normal and inflamed tissues, revealed a higher incidence of endotoxin-induced thrombosis and displayed deficient plasmin-mediated macrophage function. Mice with combined deficiency of t-PA and u-PA survived embryonic development but suffered retarded postnatal growth, markedly reduced fertility and shortened life span. Furthermore, spontaneous fibrin deposition occurred more extensively, at an earlier age and in more organs than in mice with single u-PA deficiency.

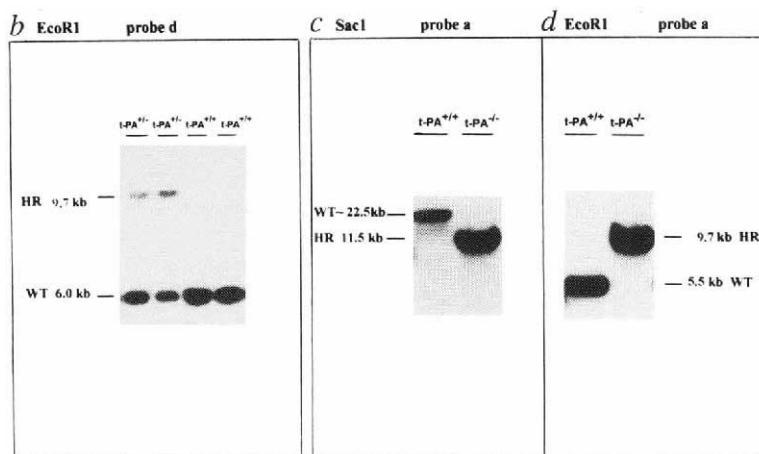
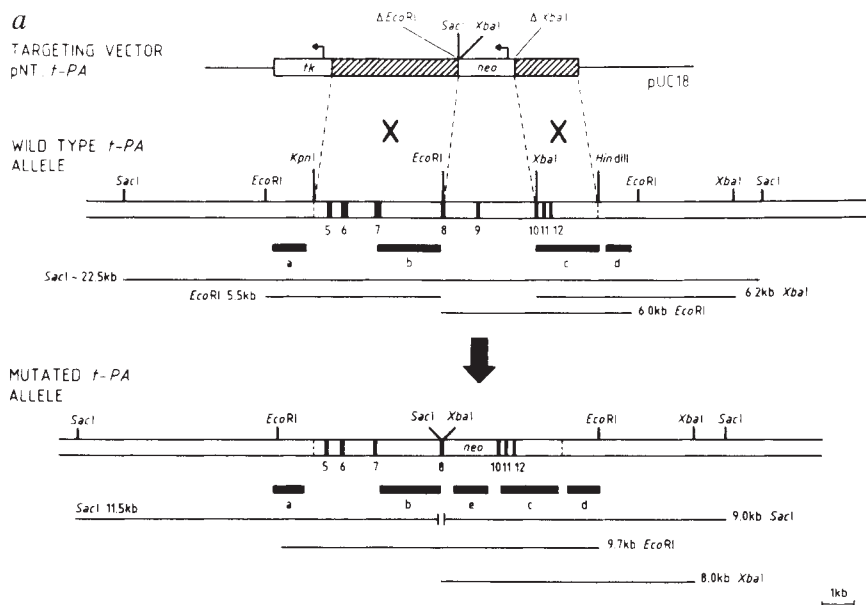
Targeting of the t-PA and u-PA genes

To disrupt the t-PA gene, the targeting vector pNT.t-PA was constructed in which a cassette expressing the neomycin-resistance gene (*neo*) replaced the genomic sequences encoding most of the kringle-2 domain and part of the proteinase domain, including the catalytically essential histidine residue at position 326 (Fig. 1a). In the targeting vector used to disrupt the u-PA gene, pNT.u-PA, *neo* replaced the genomic sequences encompassing all but 23 amino acids of the coding sequence (Fig. 2a). Thus, the correct homologous recombination resulted in total

† Present address: Center for Molecular and Vascular Biology, University of Leuven, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

To whom correspondence should be addressed.

FIG. 1. Outline of the strategy used to disrupt the t-PA gene, and Southern blot analysis of genomic DNA. **a**, Schematic representation of the targeting vector pNT.t-PA used for disruption of the t-PA gene, of the wild-type t-PA allele and of the homologously recombined t-PA allele. pNT.3't-PA was generated by cloning a 2.0-kilobase (kb) 3'-flanking XbaI-HindIII fragment of the murine t-PA gene (isolated from a D₃ ES cell genomic DNA library provided by R. Jaenisch) into the XbaI site of the parental vector pNT by blunt-end ligation³⁴. pNT.t-PA was constructed by blunt-end ligation of a 4.3 kb 5'-flanking KpnI-EcoRI fragment of the murine t-PA gene into the EcoRI site of pNT.3't-PA. The EcoRI site of the 5'-flanking KpnI-EcoRI fragment is at base pair (bp) 810 and the XbaI site of the 3'-flanking XbaI-HindIII fragment is at bp 1,170 of the murine t-PA cDNA (notation as in ref. 35). The downstream EcoRI site in the 5'-flanking fragment and the upstream XbaI site in the 3'-flanking fragments were destroyed during the cloning process as denoted by Δ EcoRI and Δ XbaI. The targeting vector, wild-type and mutated allele are represented by double lines; shaded boxes denote 5'- and 3'-flanking regions of the targeting vector; filled boxes denote exons. Relative sizes of introns are approximate and based on mapping information of the rat t-PA gene³⁶. Only exons that were mapped by Southern blot analysis are shown and numbered. The expected restriction fragments of the wild-type and the targeted mutated allele are indicated with their relative sizes by underlining. Black boxes under the genomic structure denote the probes used for Southern blot analysis. Probe 'a' is a 0.5 kb EcoRI-XbaI genomic DNA fragment, located adjacent at the 5' site of the flanking regions of pNT.t-PA. Probe 'b' is a 2.2-kb XbaI-EcoRI genomic DNA fragment and probe 'c' a 2.0-kb XbaI-HindIII genomic DNA fragment, contained in the 5' or 3' flanking regions, respectively. Probe 'd' encompasses a 1.0-kb HindIII-EcoRI genomic fragment adjacent to the 3' XbaI-HindIII fragment used for the targeting construct. Probe 'e' is a 600-bp PstI fragment from the neo gene. **b**, Southern blot analysis of genomic DNA from D₃ embryonic stem cells, digested with EcoRI and hybridized to the 3' t-PA flanking probe 'd', showing two t-PA^{+/+} ES cell colonies, harbouring a 6.0-kb EcoRI fragment of the wild-type allele and a 9.7-kb EcoRI fragment of the mutated t-PA allele. t-PA^{+/+} ES cell clones were identified in 10% of all double-resistant ES cell clones, electroporated with pNT.t-PA. **c**, **d**, Southern blot analysis of genomic tail-tip DNA of t-PA^{+/+}



and t-PA^{-/-} littermates, digested with SacI and EcoRI, and hybridized to probe 'a'. The observed sizes of the restriction fragments correspond to those expected for homologous recombination of the targeting vector pNT.t-PA into the t-PA locus.

METHODS. Culture, transfection and selection of embryonic stem cells, and isolation, restriction digestion and Southern blot analysis of genomic DNA were as previously described³⁷.

inactivation of the t-PA and u-PA genes, respectively. Site-specific and single homologous recombination at the t-PA and u-PA locus, respectively, were verified by Southern blot analysis of genomic DNA, using external and internal t-PA or u-PA probes, respectively, and a neo-specific probe (Figs 1 and 2).

Generation of mice with deficiencies

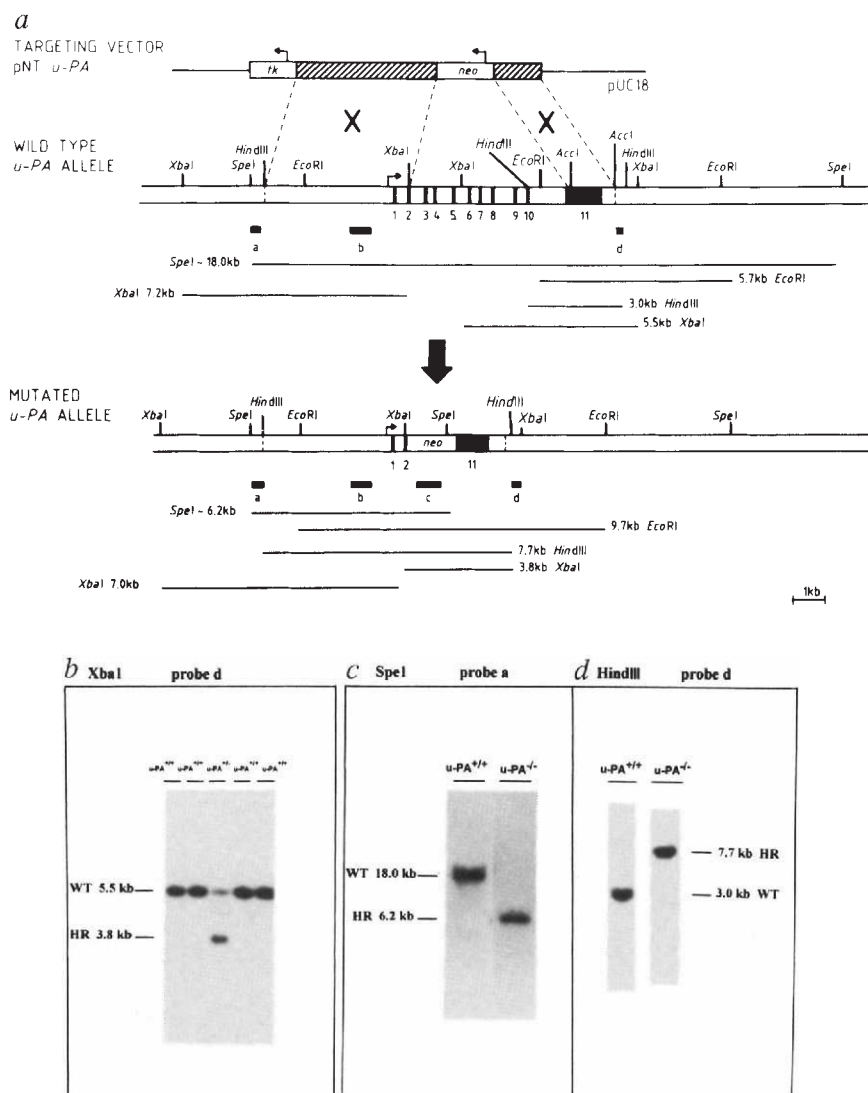
Mice with a single deficiency of t-PA or u-PA were generated by blastocyst injection of embryonic stem cell clones, harbouring a homologously recombined t-PA allele (heterozygous t-PA-deficient cells, t-PA^{+/+}; Fig. 1b) or u-PA allele (heterozygous u-PA-deficient cells, u-PA^{+/+}; Fig. 2b), respectively. Genotyping of 1,610 progeny from t-PA^{+/+} and of 703 offspring from u-PA^{+/+} breeding pairs by Southern blot analysis of tail DNA, obtained at three weeks of age (Figs 1c, d and 2c, d), revealed independent autosomal mendelian inheritance of the mutated t-PA and u-PA alleles. No differences in the litter size, the frequency of litters, the body weight of mice at 5 weeks of age or in the lifespan were found between t-PA wild-type (t-PA^{+/+})

and homozygous t-PA-deficient (t-PA^{-/-}) mice and between u-PA wild-type (u-PA^{+/+}) and homozygous u-PA-deficient (u-PA^{-/-}) mice (see supplementary information).

Correct targeting of the t-PA and u-PA genes was confirmed by the absence of detectable t-PA-specific or u-PA-specific messenger RNA and immunoreactivity (see supplementary information) and functional activity (Table 1). The t-PA activity in lung extracts and u-PA activity in kidney extracts and urine correlated closely with the numbers of functional t-PA or u-PA alleles, respectively. A compensatory increase of t-PA activity in brain extracts from u-PA^{-/-} mice or of u-PA activity in 24-h urine samples from t-PA^{-/-} mice could not be documented.

To examine the effect of a combined deficiency of t-PA and u-PA on viability, genotypes of 489 progeny from double heterozygous breeding pairs (t-PA^{+/+}:u-PA^{+/+}) were determined by Southern blot analysis at 3 weeks of age. Independent autosomal mendelian inheritance of the mutated t-PA and u-PA alleles was observed (see supplementary information). Compared to wild-type mice or mice with single deficiencies, mice with combined

FIG. 2 Outline of the strategy to disrupt the u-PA gene, and Southern blot analysis of genomic DNA. **a**, Schematic representation of the targeting vector pNT.u-PA used for disruption of the u-PA gene, of the wild-type u-PA allele and of the homologously recombined u-PA allele. pNT.5'-u-PA was generated by cloning a 4.3-kb fragment from plasmid pBSMuPAneo[8.60riB]tk (provided by J. Degen), containing the 5'-flanking *HindIII*-*XbaI* genomic fragment, into the *XbaI* site of pNT. The *XbaI* site in the 5'-flanking fragment is at bp +398 of the murine u-PA gene, 22 bp upstream of the ATG start codon (notation as in ref. 38). A 1.3-kb 3'-flanking *AccI* genomic fragment (isolated from the D₃ ES cell genomic DNA library, provided by R. Jaenisch) was then cloned into the *XhoI* site of pNT.5'-u-PA by blunt end ligation to form pNT.u-PA. The upstream *AccI* in the 3'-flanking fragment is at bp +5,711 of the murine u-PA gene, 72 bp upstream of TGA stop codon³⁸. The resulting targeted u-PA allele contains exon 1 and 14 bp of exon 2 (non-coding sequence) and 999 bp of exon 11 (including 23 amino acids of coding sequence). Notation is as in Fig. 1a. Probe 'a' is a 0.20-kb *SpeI*-*HindIII* genomic fragment, adjacent 5' to the *HindIII*-*XbaI* flanking region; probe 'b' is a 0.64-kb *PvuII* genomic fragment contained in the 5'-flanking fragment (bp -963 to -320 of the murine u-PA gene)³⁸; probe 'c' is a 0.60-kb *PstI* fragment from the *neo* gene; probe 'd' is a 0.24-kb *AccI*-*HindIII* genomic fragment (bp +7,040 to +7,280 of the murine u-PA gene)³⁸, adjacent 3' to the *AccI* flanking fragment. **b**, Southern blot analysis of genomic DNA from D₃ embryonic stem cells, digested with *XbaI* and hybridized to the 3' u-PA flanking probe 'd', revealing an ES cell colony, harbouring a 5.5-kb *XbaI* fragment of the wild-type allele and a 3.8-kb *XbaI* fragment of the mutated u-PA allele. u-PA^{+/+} ES cell clones were identified in 2.5% of all double resistant ES cell clones, electroporated with pNT.u-PA. **c, d**, Southern blot analysis of genomic tail-tip DNA of u-PA^{+/+} and u-PA^{-/-} littermates, digested with *SpeI* and *HindIII*, and hybridized to probe 'a' and 'd', respectively. The observed sizes of the restriction fragments correspond to those expected for homologous recombination of the targeting vector pNT.u-PA into the u-PA locus.



METHODS. Culture, transfection and selection of embryonic stem cells, and isolation, restriction digestion and Southern blot analysis of genomic DNA were as previously described³⁸.

homozygous deficiency of t-PA and u-PA (t-PA^{-/-}:u-PA^{-/-}) had a shorter lifespan (9 of 22 t-PA^{-/-}:u-PA^{-/-} mice but all of 15 wild-type mice survived until the end of an observation period of 30 to 40 weeks, whereas 3 of 10 double-deficient and all of 25 wild-type mice survived for 40 to 50 weeks; $P < 0.001$, by chi-square analysis). Double-deficient mice suffered growth retardation beyond the age of 3 weeks (body weights were 22 ± 1.5 g, $n = 14$, for t-PA^{-/-}:u-PA^{-/-} mice versus 33 ± 2.2 g, $n = 11$, for t-PA^{+/+}:u-PA^{+/+} mice at 23 weeks, $P < 0.0001$). They were also significantly less fertile: of the 19 double-deficient breeding pairs that were test-bred for at least 10 weeks, 6 did not produce any litter, 4 produced 1 litter, 7 produced 2 litters and 2 produced 3 litters. In contrast, 6 of 9 wild-type breeding pairs produced 2 litters and 3 of 9 wild-type breeding pairs produced 3 litters during similar breeding periods (see supplementary information). t-PA^{-/-}:u-PA^{-/-} breeding pairs produced slightly fewer offspring per litter than t-PA^{+/+}:u-PA^{+/+} breeding pairs: 4.8 ± 0.4 ($n = 30$) versus 6.2 ± 0.3 mice per litter ($n = 53$; $P = 0.007$).

Macroscopic and microscopic analysis

No macroscopic abnormalities were observed in t-PA^{-/-} mice up to an age of 14 months. u-PA^{-/-} mice developed rectal pro-

lapse of a non-infectious origin (6 out of 281 mice of 22 ± 0.8 weeks of age; 9%) and/or suffered extensive non-healing ulcerations at the eyelids, the ears around the eartag and the face, possibly resulting from the tagging trauma or scratching against the cage grid during feeding (12 of 281 mice of 26 ± 4 weeks of age; 5%).

From the age of 8 to 12 weeks on, t-PA^{-/-}:u-PA^{-/-} mice developed the following pathological conditions with variable penetrance and onset. Rectal prolapse developed in t-PA^{-/-}:u-PA^{-/-} mice at a significantly higher incidence than in u-PA^{-/-} mice (22 of 59 mice; 37%, $P < 0.001$ versus u-PA^{-/-} mice) and at an earlier age (12 ± 0.8 weeks; $P < 0.001$). As observed in u-PA^{-/-} mice, a small percentage of t-PA^{-/-}:u-PA^{-/-} mice (3 of 59 of 24 ± 1 weeks; 5%) had extensive non-healing ulcerations. A significant number of t-PA^{-/-}:u-PA^{-/-} mice (10 of 59; 17%) became runted, cachectic, dyspneic and died at 17 ± 2 weeks of age (the microscopical analysis of 6 of those is presented below). Such severe illness and early death was not observed in wild-type mice nor in mice with single inactivation of the t-PA or u-PA genes.

No histological abnormalities could be documented in any of the t-PA^{-/-} mice. u-PA^{-/-} mice displayed occasional occurrence of small, focal fibrin deposits in the intestines and in the sinusoids

FIG. 3 Light microscopic analysis of liver, colon, uterus and lung of preterminal t-PA^{-/-}:u-PA^{-/-} mice (3 males of 8 to 15 weeks; 3 females of 12 to 23 weeks). **a**, Liver section revealing amorphous fibrin deposition (arrow) under the liver capsula (observed in all mice; haematoxylin–eosin staining; magnification, $\times 195$). Calcification (asterisk) of the fibrin deposits was frequently observed. **b**, Adjacent liver section stained according to von Kossa revealing black calcified lesions throughout the subcapsular liver parenchyma (magnification, $\times 79$). **c**, Immunostaining of a liver section with a murine fibrinogen/fibrin-specific antiserum (magnification, $\times 79$), revealing fibrin deposit. Specificity of the immunohistochemical staining was confirmed by preadsorption of the antiserum with murine fibrinogen which abolished positive staining, by the absence of similar fibrin-immunoreactive deposits in liver sections of wild-type mice and by the lack of staining with irrelevant primary antisera (data not shown). **d, e**, Colon sections stained with haematoxylin–eosin, revealing a superficial epithelial ulceration covered with fibrin (arrows in **d**; magnification, $\times 79$) and a subepithelial fibrin deposition (arrow in **e**; magnification, $\times 195$) (observed in 4 of 5 preterminal t-PA^{-/-}:u-PA^{-/-} mice). **f**, Uterine section stained with haematoxylin–eosin revealing a large, subepithelial fibrin deposition (arrow) which extends in the stroma (asterisk) of the uterus (observed in 2 of 3 examined t-PA^{-/-}:u-PA^{-/-} female mice; magnification, $\times 195$). **g**, Lung section stained with haematoxylin–eosin, revealing a triangular collapsed region of lung parenchyma, containing abundant fibrin deposition, which stained specifically with the murine fibrinogen/fibrin-specific antiserum (data not shown) (observed in 5 of 6 examined t-PA^{-/-}:u-PA^{-/-} preterminal mice; magnification, $\times 79$).

METHODS. Preparation, fixation and staining of the tissues was as previously described³⁷. For immunohistochemistry, 5- μ m sections from snap frozen liver and lung tissue were stained with a three-step indirect immunoperoxidase-method using primary goat anti-mouse fibrinogen/fibrin antibody, followed by peroxidase-conjugated rabbit anti-goat, and swine anti-rabbit immunoglobulins as second and third layers, respectively. Each incubation was at room temperature and followed by a 15-min wash in three changes of PBS, pH 7.2. The reaction product was developed by incubation for 15 min in 0.05 M acetate buffer (pH 4.9) containing

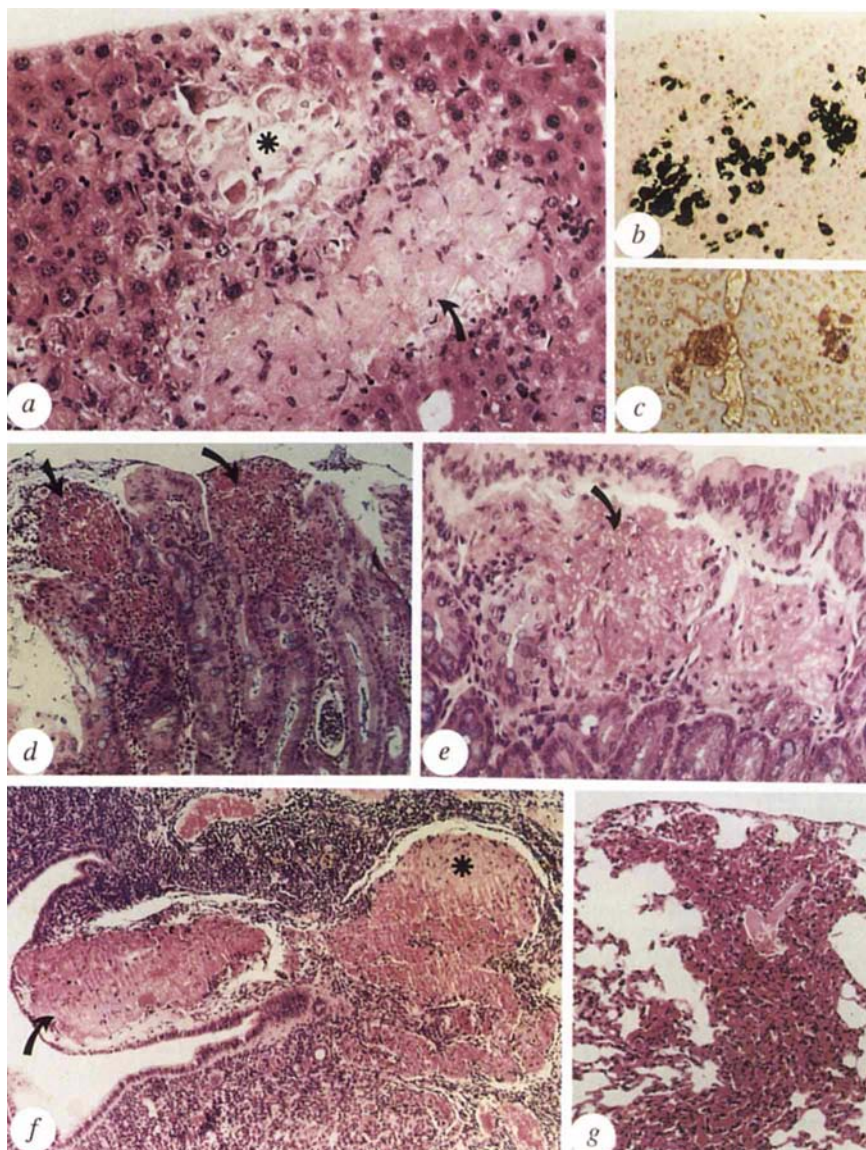
0.05% 3-amino-9-ethylcarbazole and 0.01% H₂O₂, resulting in a red staining of immunoreactive sites. The sections were briefly counterstained with Harris' haematoxylin, and mounted with glycerin jelly.

of the liver and excessive fibrin deposits in the ulcerated skin, ear or prolapsed rectum (not shown). Histological abnormalities were only documented in t-PA^{-/-}:u-PA^{-/-} mice beyond 2 to 3 months of age, with variable onset and penetrance and most prominently in preterminal t-PA^{-/-}:u-PA^{-/-} mice. Preterminal t-PA^{-/-}:u-PA^{-/-} mice revealed spontaneous fibrin deposits in liver, intestines, gonads, lungs (Fig. 3), in non-healing ulcerative lesions of the skin, ears and prolapsed rectum and occasionally in the kidneys (not shown). The fibrinous nature of the deposits was confirmed by immunostaining of frozen sections with a goat fibrinogen/fibrin-specific antiserum (Fig. 3c) and by electron-microscope analysis, which revealed the typical fibrillar structure and periodicity of fibrin polymers (not shown). Intestinal adhesions and occasionally ischaemic tissue necrosis (uterus and intestines), possibly resulting from thrombosis, were observed (not shown). The incidence of fibrin deposition in liver, intestines, gonads and lung was significantly higher in t-PA^{-/-}:u-PA^{-/-} mice than in mice with any of the other genotypes studied

($P < 0.001$ to $P < 0.05$) (see supplementary information).

Thrombolysis and thrombosis

The role of t-PA and u-PA in spontaneous clot lysis was evaluated by determining the *ex vivo* lysis of ¹²⁵I-fibrin-labelled plasma clots, which were injected through the jugular vein and embolized into the pulmonary arteries²⁵. No difference in the rate of spontaneous plasma clot lysis was observed between the wild-type and u-PA^{-/-} mice ($85 \pm 4\%$ versus $83 \pm 9\%$ lysis over 24 h, respectively; P not significant) (Fig. 4). In contrast, t-PA^{-/-} mice lysed pulmonary plasma clots at a significantly reduced rate ($21 \pm 3\%$ lysis over 24 h, $P < 0.001$ versus wild-type mice) but pulmonary clot lysis was most prominently reduced in the t-PA^{-/-}:u-PA^{-/-} mice ($3 \pm 3\%$ lysis within 24 h in t-PA^{-/-}:u-PA^{-/-} mice, $P < 0.002$ versus t-PA^{-/-} mice). Pulmonary plasma clot lysis remained significantly lower in the combined t-PA^{-/-}:u-PA^{-/-} mice than in the t-PA^{-/-} mice for up to 48 h, but was not significantly different after 72 h (Fig. 4).



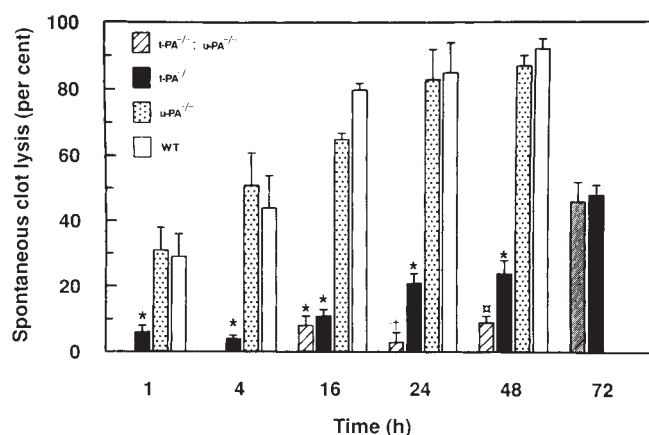


FIG. 4 Spontaneous lysis of a ^{125}I -fibrin-labelled pulmonary plasma clot. A $25\ \mu\text{l}$ ^{125}I -fibrin-labelled plasma clot was injected into the jugular vein and the residual radioactivity in heart and lungs measured *ex vivo* at the indicated time points. Values are expressed as a per cent of total radioactivity injected. $t\text{-PA}^{-/-}$ mice lysed ^{125}I -fibrin-labelled plasma clots more slowly than $u\text{-PA}^{-/-}$ or wild-type mice; lysis of the plasma clots was slowest in $t\text{-PA}^{-/-}; u\text{-PA}^{-/-}$ mice. *, $P < 0.001$ versus wild-type and versus $u\text{-PA}^{-/-}$ mice. †, $P < 0.002$, ○, $P = 0.01$ versus $t\text{-PA}^{-/-}$ mice. Preparation *in vitro* and lysis *in vivo* of ^{125}I -fibrin-labelled plasma clots in mice was as previously described.²⁵

$t\text{-PA}^{-/-}$ and $u\text{-PA}^{-/-}$ mice were tested for their susceptibility to venous thrombosis after local injection of proinflammatory endotoxin in the footpad, a model which is sensitive to alterations of the fibrinolytic system²⁶. Injection of 10 or 20 μg endotoxin (endotoxin W. from *Escherichia coli* 0111:B4 or from *Salmonella typhosa* 908) induced venous thrombosis in 29 of 54 wild-type mice (54%), in 29 out of 38 $t\text{-PA}^{-/-}$ mice (76%, $P = 0.046$ versus wild-type mice, by chi-square analysis) and in 27 out of 30 $u\text{-PA}^{-/-}$ mice (90%, $P = 0.002$ versus wild-type mice). The extent of venous thrombosis, as estimated from the number of veins with visible thrombi, was significantly higher in the $t\text{-PA}^{-/-}$ and $u\text{-PA}^{-/-}$ mice. Only 15% of the wild-type but 55% of the $t\text{-PA}^{-/-}$ mice ($P < 0.001$) and 60% of the $u\text{-PA}^{-/-}$ mice ($P < 0.001$) contained more than 4 thrombosed veins per analysed tissue section.

Macrophage function

To examine whether $u\text{-PA}$ might allow extracellular matrix degradation^{8,27}, lysis of ^{125}I -fibrin-labelled matrix and of ^3H -proline-labelled subendothelial matrix by thioglycollate-stimulated macrophages (which have a 10- to 100-fold enhanced production of $u\text{-PA}$) was studied. Disruption of the $u\text{-PA}$ genes specifically abolished plasminogen-dependent breakdown of ^{125}I -fibrin-labelled matrix and of ^3H -proline-labelled subendothelial matrix breakdown by thioglycollate-activated macrophages (Fig. 5), but $u\text{-PA}$ deficiency did not affect invasion of macrophages into the peritoneal cavity after thioglycollate injection (Fig. 5). Matrix degradation by thioglycollate-activated macrophages was not affected by inactivation of the $t\text{-PA}$ genes (Fig. 5). These results indicate that $u\text{-PA}$ -mediated plasmin formation may aid in breakdown of fibrin deposits or matrix components, but that in the absence of $u\text{-PA}$, invasion by macrophages is unaltered under the present experimental conditions.

Discussion

The physiological plasminogen activators $t\text{-PA}$ and $u\text{-PA}$ share the characteristic ability to activate plasminogen, but have distinct protein structures, tissue-specific expression and biological activities, suggesting that these molecules may play distinct roles in different biological processes¹⁻³. Much circumstantial evidence

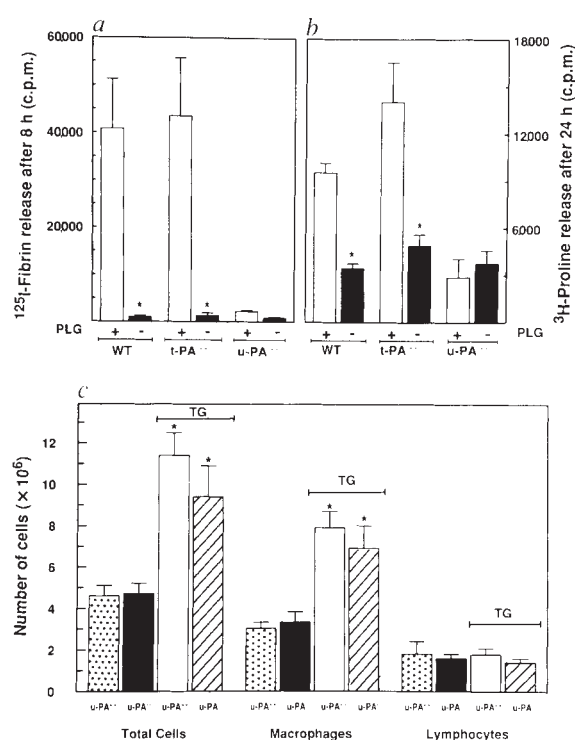


FIG. 5 Matrix degradation and invasion into the peritoneal cavity by thioglycollate-stimulated macrophages. a, b, Lysis of ^{125}I -fibrin matrix (a) and of ^3H -proline-labelled subendothelial matrix (b) by thioglycollate-stimulated macrophages from wild-type, $t\text{-PA}^{-/-}$ and $u\text{-PA}^{-/-}$ mice. PLG (+): with $16\ \mu\text{g}\ \text{ml}^{-1}$ plasminogen in the culture medium; PLG (-): without plasminogen; WT: wild-type. The data represent mean $\pm$ s.e.m. of triplicate measurements of 5 or 6 independent experiments. $P < 0.05$ versus PLG (+). c, Invasion into the peritoneal cavity of thioglycollate-stimulated macrophages from $u\text{-PA}^{+/+}$ and $u\text{-PA}^{-/-}$ mice. Macrophages from $u\text{-PA}^{-/-}$ mice invaded the peritoneal cavity equally well as macrophages from $u\text{-PA}^{+/+}$ mice, both with or without thioglycollate injection. The data represent mean $\pm$ s.e.m. of 6 experiments. Cell counts were done on at least 200 cells. TG, with intraperitoneal injection of 0.5 ml thioglycollate; *, $P < 0.05$ versus control. METHODS. Preparation and lysis of ^{125}I -fibrin and ^3H -proline subendothelial matrix by thioglycollate-stimulated peritoneal macrophages was as described¹⁰. Invasion was evaluated by injecting mice intraperitoneally with 0.5 ml thioglycollate and counting, after 3 days, the number of different cell types in the peritoneal lavage⁹.

implicates these plasminogen activators in ovulation^{11,12}, sperm migration and fertilization¹⁶, embryo implantation^{17,20}, embryogenesis^{18,19} and associated tissue remodelling of the ovary^{17,28}, prostate, and mammary gland^{29,30}, but no genetic deficiencies involving either gene product have been described in man. Thus, inactivation of $t\text{-PA}$ or $u\text{-PA}$ genes in mice might have been anticipated to result in a lethal phenotype. Surprisingly, single- and double-deficient mice appear normal at birth, suggesting that neither $t\text{-PA}$ nor $u\text{-PA}$, individually or in combination, is required for normal embryonic development. Although doubly deficient mice are able to reproduce, they are significantly less fertile than wild-type mice or mice with a single deficiency of $t\text{-PA}$ or $u\text{-PA}$, possibly because of the poor general health of these animals (low body weight, dyspnoea, rectal prolapse and cachexia) or to the large fibrin deposits in their gonads. These data suggest that proteinases³¹ other than $t\text{-PA}$ and $u\text{-PA}$ may be more essential in reproduction and embryonic development than previously suspected.

These targeting experiments reveal a divergent and specific role for each plasminogen activator in fibrinolysis and macrophage function. The role of $t\text{-PA}$ and $u\text{-PA}$ in the normal surv-

TABLE 1 t-PA and u-PA activity in wild-type and homozygous t-PA-deficient (t-PA^{-/-}) or u-PA-deficient (u-PA^{-/-}) mice

Mice	t-PA activity		u-PA activity	
	Lung (IU mg ⁻¹)	Brain (IU mg ⁻¹)	Kidney (IU mg ⁻¹)	Urine (IU)
t-PA ^{+/+}	5.0 ± 0.8 (5)	5.6 ± 0.4 (5)	ND	99 ± 35 (8)
t-PA ^{+/-}	1.7 ± 0.3* (5)	ND	ND	ND
t-PA ^{-/-}	<0.1 (5)	<0.3 (5)	ND	120 ± 23 (9)
u-PA ^{+/+}	ND	5.6 ± 0.8 (5)	0.46 ± 0.1 (5)	131 ± 19 (9)
u-PA ^{+/-}	ND	ND	0.22 ± 0.02* (5)	ND
u-PA ^{-/-}	ND	5.7 ± 1.3 (3)	<0.05 (5)	<3.1

Values represent mean ± s.e.m. of activities from the number of animals indicated in brackets. t-PA and u-PA activities are expressed as IU of murine t-PA or murine u-PA per mg protein in the tissue extracts (lung, brain, kidney) or as total units in 24-h urine samples. t-PA and u-PA activities were determined by the fibrin plate method³² (lung and kidney) or by a chromogenic substrate assay³³ (brain and urine) in the presence of 1 mM amiloride (which blocks murine u-PA) or of 200 µg ml⁻¹ murine t-PA-specific IgGs, respectively. The specific activity of murine u-PA, purified from endothelioma-conditioned medium¹³, was 26,000 IU mg⁻¹ and that of recombinant murine t-PA was 475,000 IU mg⁻¹, by comparison with the human standards, obtained from the National Institute for Biological Standards and Control (Hertford, UK). ND, not determined.

**P* < 0.05 versus wild-type mice. A direct correlation is observed between t-PA or u-PA activity in different tissues and the number of functional t-PA or u-PA alleles, respectively. No compensatory increase of t-PA or u-PA activity is observed in u-PA- or t-PA-deficient mice, respectively. t-PA or u-PA activities in tissue extracts from t-PA^{+/+} or u-PA^{+/+} mice, respectively, were not reduced by addition of tissue extracts from t-PA^{-/-} or u-PA^{-/-} mice, respectively (data not shown), indicating that absence of detectable t-PA or u-PA activity in tissue extracts from t-PA^{-/-} or u-PA^{-/-} mice, respectively, did not result from increased proteinase inhibitor activity.

eillance of spontaneous fibrin deposition is supported by the observation that t-PA^{-/-} and t-PA^{-/-}:u-PA^{-/-} mice displayed a reduced or virtually absent endogenous lysis of ¹²⁵I-fibrin-labelled plasma clots, respectively, and the finding that u-PA-deficient and doubly deficient mice suffered occasional or extensive spontaneous fibrin deposition, respectively. In addition, in response to proinflammatory endotoxin injection *in vivo*, overt thrombosis was observed in both t-PA- and u-PA-deficient mice at a significantly increased incidence and to a larger extent than

in wild-type mice. Possibly, impaired fibrin dissolution by u-PA-deficient macrophages may constitute a mechanism for the increased occurrence of venous thrombosis after endotoxin injection and the abnormal fibrin deposition in liver, skin and intestine in u-PA^{-/-} and t-PA^{-/-}:u-PA^{-/-} mice. Collectively, these data confirm the role of the fibrinolytic system in maintaining vascular patency² and suggest that fibrinolysis may play a significant role in preventing thrombosis in conditions of inflammation and injury. Further examination of the effect of t-PA and u-PA gene disruption in pathological disorders such as atherosclerosis²², glomerulonephritis²³ and malignancy¹⁵ seems to be warranted.

One unresolved issue is whether the observed phenotype of spontaneous fibrin deposition in t-PA^{-/-}:u-PA^{-/-} mice occurs as a result of the markedly reduced thrombolytic potential or secondary to deficient tissue remodelling. u-PA has been proposed to modulate tissue remodelling through activation of matrix-degrading proteinases and growth factors^{6,7}. Finally, the apparently normal invasiveness of u-PA^{-/-} macrophages and trophoblasts (as revealed by the normal reproduction of u-PA^{-/-} mice) might suggest a less prominent role of u-PA in cellular invasion than previously suspected. Other processes of cellular invasion such as in tumorigenesis¹⁵ and endothelial or vascular smooth muscle cell migration^{3,6,7} still need to be examined further.

The phenotypes of mice with a single deficiency of t-PA or u-PA would appear to result, at least in part, from the ability of one plasminogen activator to complement the other. Mice with a combined deficiency of t-PA and u-PA present with spontaneous fibrin depositions at a significantly increased incidence, to a larger extent, and in more organs than observed in the single mutants, and suffer retarded postnatal growth and a markedly reduced fertility and life expectancy. The almost total lack of endogenous lysis of a ¹²⁵I-fibrin plasma clot and the higher incidence and earlier occurrence of the rectal prolapse in the doubly deficient mice further support the notion that u-PA and t-PA cooperate in a variety of biological phenomena, possibly related to fibrin degradation, matrix remodelling and/or growth factor activation^{6,7}.

This study indicates that the t-PA and u-PA gene products have divergent roles in fibrinolysis and macrophage function and that they cooperate in a number of complex biological processes. Although their combined deficiency results in increased morbidity and mortality, it does not appear to affect embryonic development. □

Received 19 November 1993; accepted 3 February 1994.

- Astrup, T. in *Progress in Chemical Fibrinolysis and Thrombolysis* Vol. 3 (eds Davidson, J.F., Rowan, R. M., Samama, M. M. & Desnoyers, P. C.) 1–57 (Raven, New York, 1987).
- Collen, D. & Lijnen, H. R. *Blood* **78**, 3114–3124 (1991).
- Vassalli, J. D., Sappino, A. P. & Belin, D. *J. clin. Invest.* **88**, 1067–1072 (1991).
- Kluft, C. *Tissue-type Plasminogen Activator (t-PA): Physiological and Clinical Aspects* Vol. 1 (ed. Kluft, C.) 47–79 (CRC, Boca Raton, Florida, 1988).
- Blasi, F., Vassalli, J. D. & Dane, K. *J. Cell Biol.* **104**, 801–804 (1987).
- Saksela, O. & Rifkin, D. B. *Rev. Cell Biol.* **4**, 93–126 (1988).
- Moscattelli, D. & Rifkin, D. B. *Biochim. biophys. Acta* **948**, 67–85 (1988).
- Unkles, J. C., Gordon, S. & Reich, E. *J. exp. Med.* **139**, 834–850 (1974).
- Freyria, A. M., Paul, J., Belleville, J., Broyer, P. & Etoy, R. *Comp. Biochem. Physiol.* **99**, 517–524 (1991).
- Baker, M. S., Bleakley, P., Woodrow, G. C. & Doe, W. F. *Cancer Res.* **50**, 4676–4684 (1990).
- Huarte, J., Belin, D. & Vassalli, J. D. *Cell* **43**, 551–558 (1985).
- Tsafiri, A., Bicsak, T. A., Cajander, S. B., Ny, T. & Hsueh, A. J. W. *Endocrinology* **124**, 415–421 (1989).
- Wagner, E. F. *EMBO J.* **9**, 3024–3032 (1990).
- Morioka, S., Lazarus, G. S., Baird, J. L. & Jensen, P. J. *J. Invest. Dermatol.* **88**, 418–423 (1987).
- Dano, K. et al. *Adv. Cancer Res.* **44**, 139–266 (1985).
- Huarte, J., Belin, D., Bosco, D., Sappino, A. P. & Vassalli, J. D. *J. Cell Biol.* **104**, 1281–1289 (1987).
- Sappino, A. P., Huarte, J., Belin, D. & Vassalli, J. D. *J. Cell Biol.* **109**, 2471–2479 (1989).
- Strickland, S., Reich, E. & Sherman, M. I. *Cell* **9**, 231–240 (1976).
- Menoud, P. A., Debrot, S. & Schowing, J. *Roux Arch. dev. Biol.* **198**, 219–226 (1989).
- Denker, H. W. *Adv. Anat. Embryol. Cell Biol.* **53**, 3–123 (1977).
- Nilsson, I. M., Ljungner, H. & Tengborn, L. *Br. Med. J. Clin. Res. Ed.* **290**, 1453–1456 (1985).

- Juhan-Vague, I. & Colien, D. *Ann. Epidemiol.* **2**, 427–438 (1992).
- Tomooka, S., Border, W. A., Marshall, B. C. & Nobel, N. A. *Kidney Int.* **42**, 1462–1469 (1992).
- Idell, S. et al. *J. clin. Invest.* **84**, 695–705 (1989).
- Stassen, J. M., Vanlinthout, I., Lijnen, H. R. & Colien, D. *Fibrinolysis* **4** (suppl. 2) 15–21 (1990).
- Carmeliet, P. et al. *J. clin. Invest.* **92**, 2756–2760 (1993).
- Granelli-Piperno, A., Vassalli, J. D. & Reich, E. *J. exp. Med.* **146**, 1693–1706 (1977).
- Canipari, R., O'Connell, M. L., Meyer, G. & Strickland, S. *J. Cell Biol.* **105**, 977–981 (1987).
- Busso, N., Huarte, J., Vassalli, J. D., Sappino, A. P. & Belin, D. *J. biol. Chem.* **264**, 7455–7457 (1989).
- Ossowski, L., Biegel, D. & Reich, E. *Cell* **16**, 929–940 (1979).
- Lala, P. K. & Graham, C. H. *Cancer Metastasis Rev.* **9**, 369–379 (1990).
- Astrup, T. & Müllertz, S. *Arch. Biochem. Biophys.* **40**, 346–351 (1952).
- Verheijen, J. H., Chang, G. T. G. & Kluft, C. *Thromb. Haemostasis* **51**, 392–395 (1984).
- Tybuliewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
- Rickles, R. J., Darrow, A. L. & Strickland, S. *J. biol. Chem.* **263**, 1563–1569 (1988).
- Peng, F., Ohlsson, M. & Ny, T. *J. biol. Chem.* **265**, 2022–2027 (1990).
- Carmeliet, P. et al. *J. clin. Invest.* **92**, 2746–2755 (1993).
- Degen, S. J., Heckel, J. L., Reich, E. & Degen, J. L. *Biochemistry* **26**, 8270–8279 (1987).

ACKNOWLEDGEMENTS. We thank P. Holvoet and L. Nelles for cloning and expression of murine t-PA, and A. Bouché, C. Declercq, E. Demarsin, M. De Mol, S. Janssens, S. Pollefeet, J. M. Stassen and S. Wynn for technical assistance. P.C. is a research associate of the NFWO. This work was supported by the NIH (R.C.M.) and the Belgian Instituut voor Wetenschap en Technologie and the Inter-Universitaire Attractiepolen (P.C., L.S., L.K., D.C.).

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London editorial office of *Nature*.

Design and synthesis of multi-haem proteins

Dan E. Robertson^{*}, Ramy S. Farid^{*}, Christopher C. Moser^{*},
Jeffrey L. Urbauer[†], Stephen E. Mulholland^{*},
Ravindernath Pidikiti^{*}, James D. Lear[‡], A. Joshua Wand[†],
William F. DeGrado^{*‡} & P. Leslie Dutton^{*}

^{*} Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[†] Department of Biochemistry, School of Chemical Sciences, University of Illinois at Urbana-Champaign, 600 S. Matthews Street, Urbana, Illinois 61801, USA

[‡] E. I. du Pont de Nemours Co., Central Research and Development Department, PO Box 80328, Wilmington, Delaware 19880, USA

A water-soluble, 62-residue, di- α -helical peptide has been synthesized which accommodates two *bis*-histidyl haem groups. The peptide assembles into a four-helix dimer with 2-fold symmetry and four parallel haems that closely resemble native haems in their spectral and electrochemical properties, including haem-haem redox interaction. This protein is an essential intermediate in the synthesis of molecular 'maquettes', a novel class of simplified versions of the metalloproteins involved in redox catalysis and in energy conversion in respiratory and photosynthetic electron transfer.

IRON porphyrin and protein combine in nature to catalyse an array of vital bioenergetic reactions. The structures of many of the smaller single and multi-haem proteins, such as myoglobin, haemoglobin, catalase, peroxidase, cytochrome P450 and the other cytochromes (cyt), are known to atomic resolution. However, the large, hydrophobic, multi-haem structures that are associated with membranes and which promote long-range electron transfers and charge separation in respiration and photosynthesis, are less well described. Of these, only the bacterial photosynthetic reaction centre protein structure is known to atomic resolution¹. Study of the other major charge-separating complexes, notably cytochrome *bc*₁ (refs 2, 3) and cytochrome *c* oxidase^{4–6}, must at present rely on working structural models. Hence, size, cofactor complexity and the lack of structural details obscure their common action in the generation of the transmembrane electrochemical potential used to drive ATP production and cellular transport^{7,8}.

Progress with *de novo* synthesis of peptides presents the opportunity to construct chemically simple peptides that incorporate biological redox sites such as haem groups, permitting us to view the molecular engineering involved in their assembly and to establish the minimal requirements for function of their highly complex natural counterparts. Hence, these synthetic proteins would play a role analogous to the small three-dimensional models called maquettes used by sculptors and architects to appreciate various aspects of full-scale products. It is essential that these structures avoid the rather nonspecific ensemble of folds commonly found in previous synthetic peptides, and to this end it is pertinent that more native-like properties have been achieved by incorporating a metal-binding site into a four-helix protein structure^{9,10}. We attempted here to confer further order on the protein by introducing the large rigid macrocycle of haem liganded to the protein via the axial positions of the iron. We describe the design and chemical synthesis of a series of proteins that incorporate up to four haem redox sites. These appear to be well organized, to display distinct binding, spectral and electrochemical properties, and to have the potential to perform long-range electron transfer.

Protein design and synthesis

We chose the dihaem cytochrome *b* subunit of cytochrome *bc*₁

as a paradigm for the design and synthesis of the haem protein. Figure 1A shows the working model structure of the haem-binding helices of the cytochrome *b* polypeptide and haem properties drawn from many studies. The charge-separating reaction of cytochrome *bc*₁ occurs through the cytochrome *b* subunit, promoted by long-range electron transfer between two haems called *b*_L and *b*_H (refs 2, 11, 12) (the subscripts L and H refer to the relatively low and high redox potentials of haems *b*_L and *b*_H). The haems (iron protoporphyrin IX) are *bis*-histidine ligated to the same pair of parallel α -helices (helices B and D)^{3,13–15}. Despite these similarities, their redox midpoint potentials differ by 140 mV. The structural and environmental factors producing the difference are unknown, but this difference is critical in providing the driving force for electron transfer from haem *b*_L to *b*_H in the transmembrane charge separation. The helices supporting the haems consist almost entirely of hydrophobic residues (see below) and lie in the apolar lipid bilayer membrane. The histidine ligands on both helices are separated by 14 residues (about four α -helical turns), leading to an inter-haem Fe-to-Fe separation of about 20 Å. Amino-acid residues Phe 104 and Ile 205 project into the interhelical region and define one edge of each haem-binding site. The side chains of three conserved Arg residues (Arg 94, Arg 114 and Arg 193) are within van der Waals contact of the propionic acid side chains of haems *b*_H and *b*_L and seem to form another edge of the haem-binding sites.

The initial synthetic protein was constructed from a 31-residue peptide comprising a 27-residue helix possessing key features of the *b*-haem-bearing helices B and D of cytochrome *bc*₁ and with a four-residue flexible tether (Cys-Gly-Gly-Gly), allowing the formation of a disulphide bond.

Cyt *bc*₁ helix D:

-Arg-Phe-Phe-Ser-Leu-His-Tyr-Leu-Leu-Pro-Phe-Val-Ile-Ala-Ala-Leu-Val-Ala-
195 200 205 210
Ile-His-Ile-Trp-Ala-Phe-
215

Cyt *bc*₁ helix B:

-Ala-Met-Arg-Tyr-Ile-His-Ala-Asn-Gly-Ala-Ser-Leu-Phe-Phe-Leu-Ala-Val-Tyr-
95 100 105
Ile-His-Ile-Phe-Arg-Gly-
110 115

Synthetic peptide:

Cys-(Gly)₅-Glu-Leu-Trp-Lys-Leu-His-Glu-Glu-Leu-Leu-Lys-Lys-Phe-Glu-Glu-Leu-
1 5 10 15 20
Leu-Lys-Leu-His-Glu-Arg-Leu-Lys-Lys-Leu-
25 30

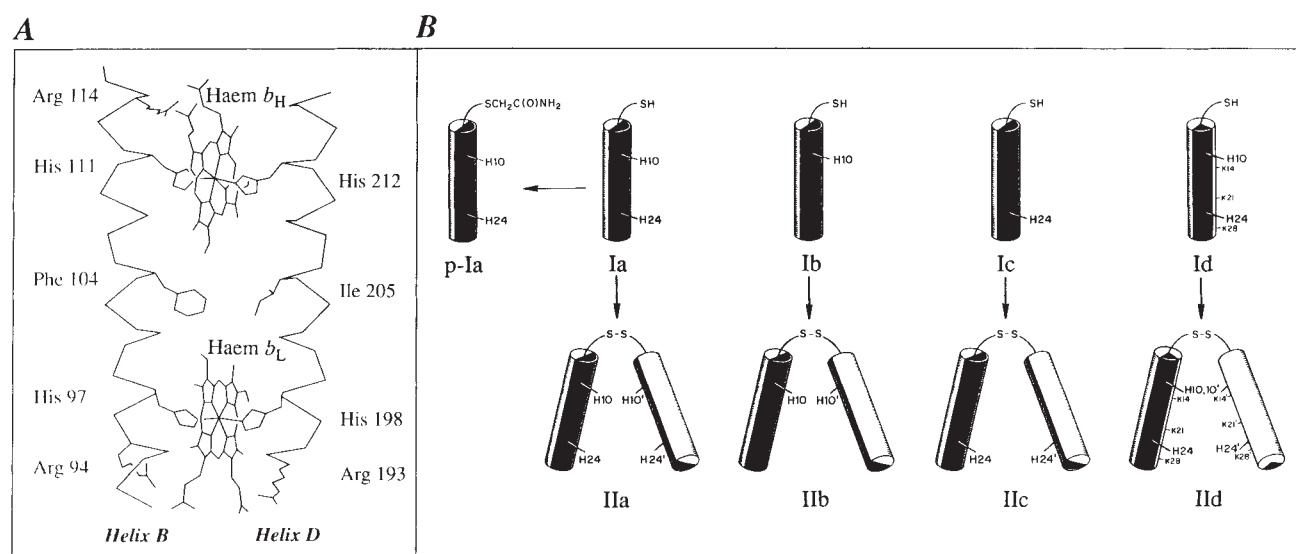


FIG. 1 A, Model of the dihelix, dihaem charge-separating unit of cytochrome bc_1 . Residue numbering is for the purple, non-sulphur photo-synthetic bacterial *Rhodobacter capsulatus* cytochrome b . Thus, haem b_L is ligated to H97/H198 and haem b_H to H111/H212 (ref. 15). The subscripts L and H refer to the relatively low and high redox potentials of haems b_L and b_H ; midpoint potentials at pH 7 are +50 and -90 mV, respectively^{21,39}. Details of the modelling will be published elsewhere (R.S.F., P.L.D. and D.E.R., manuscript in preparation). In the model the imidazole planes of the ligating histidines, in which the $N\delta$ hydrogen of each imidazole is hydrogen-bonded to the backbone carbonyl of the $i-4$ residue, adopt dihedral angles (τ) of 80° and 55° for haem b_L and b_H , respectively. These angles (see refs 24–26) are consistent with EPR spectra of cytochrome b where cyt b_L and b_H display unusual rhombic spectra with g_z values of 3.44 and 3.78, respectively⁴⁰. B, Schematic representations of the synthesized peptides. Dark shading represents hydrophobic surfaces and white the charged or polar surface of the peptide. Ia, The 31-residue bis-histidine peptide. Peptide association and haem binding were measured after blocking the cysteine thiol at

position 1 to prevent oxidation to a II-type peptide. It was protected by treatment with a 10-fold excess of iodoacetamide at pH 8.5, resulting in peptide p-Ia. Purification was done using reverse-phase HPLC. IIa was prepared from the unprotected 31-residue peptide by air oxidation at ambient temperature in 500 mM Tris buffer, pH 8.9. Monitoring of the oxidation with dithionitrobenzoic acid showed the reaction to be complete after ~4 h. Ib, Replacement of His by Ala at position 24 in Ia. After oxidation to IIb the resultant peptide was able to bind one haem per molecule at the 10,10' position. Ic, Replacement of His by Ala at position 10 in Ia. After oxidation to IIc the resultant peptide was able to bind one haem per molecule at the 24,24' position. Id, Replacement of Leu by Lys at positions 14, 21 and 28 in Ia. The modification was intended to increase hydrophilicity and decrease aggregation of IId. The I-type peptides were synthesized by standard fmoc (9-fluorenyl methyloxycarbonyl) chemistry⁴¹ and purified by reverse-phase HPLC using a C-18 column. The identity of the peptides was verified by electrospray mass spectroscopy.

The two His residues on helices B and D were incorporated at analogous positions 10 and 24 on the 31-residue peptide. The Phe on helix B was inserted at position 17 between the His residues to separate the haem-binding domains. In addition, the highly conserved Arg from the B helix was added at position 27 for the dual purpose of defining the haem-binding site at His 24 and (in principle) raising the redox potential of the haem through charge interaction^{16–18}. The remaining amino acids were generally assigned to confer amphiphilic character and water solubility to the α -helix following established principles¹⁹. Thus, the face of the helix designed to be exposed to water consisted of the polar and charged residues Glu and Lys. The other face, which included the conserved His and Phe side chains, had many apolar Leu residues, chosen to stabilize helix/helix and helix/haem interactions hydrophobically. The number of Leu residues necessary to stabilize the interactions is critical, as too few might not provide a strong enough driving force for the desired assemblies while too many could lead to the formation of non-specific aggregates and poor water solubility. In the initial design shown above, Leu occupies about one-third of the positions (9 of 31) and this peptide was expected to form two- or four-helical structures²⁰. A second design replaced three Leu residues by Lys, to show the structural effect of changing the Leu/Lys ratio. Lastly, a Trp residue was added at position 7 near the amino terminus to determine the peptide concentration ($\epsilon_{280} = 5,700 \text{ mol}^{-1} \text{ cm}^{-1}$). These peptides, and two single-His analogues, were prepared by solid-phase methods (Fig. 1B).

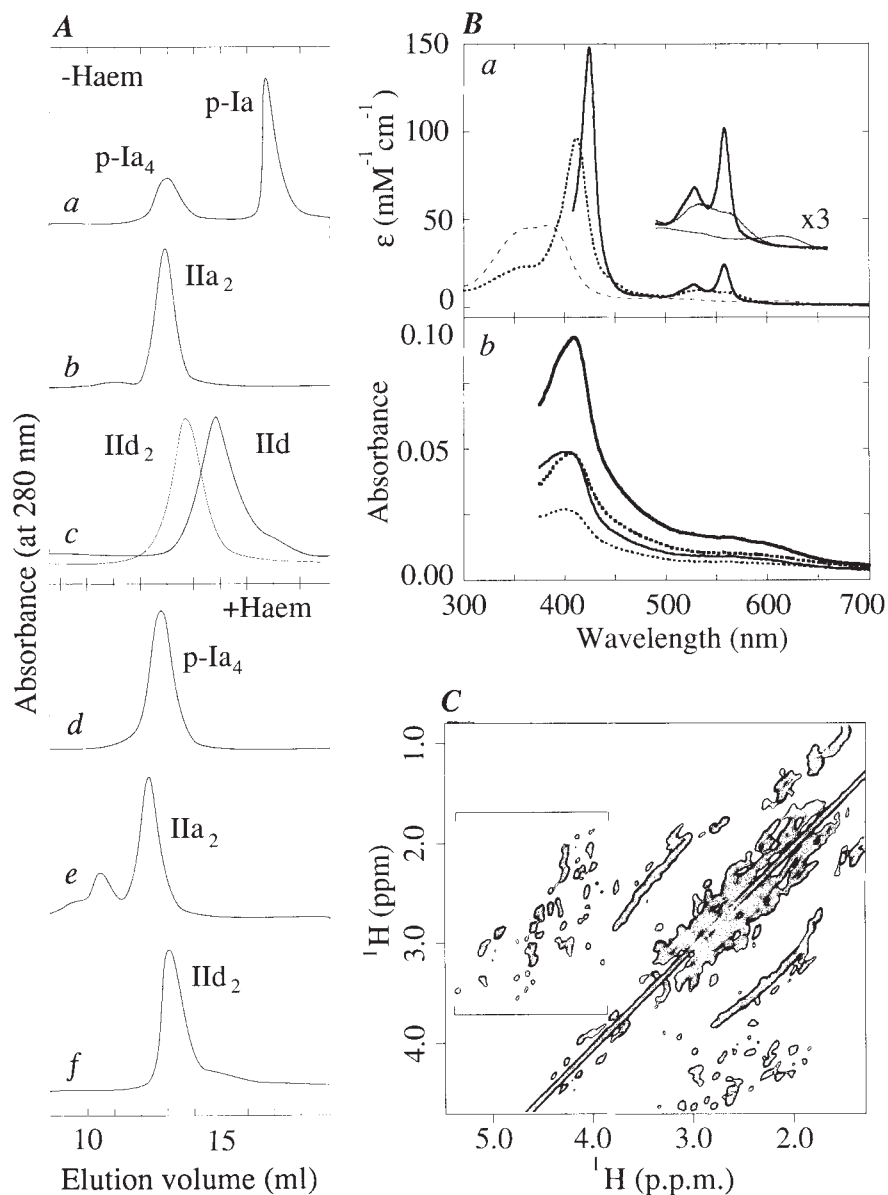
Assembly

After oxidation of the Cys thiol (Fig. 1B), all the designed pep-

tides appear to associate into assemblies consisting of a four-helix unit. Sedimentation equilibrium over a wide range of concentrations, both in the presence and absence of haem, indicates that the di- α -helical peptides IIa and IId are dimeric (Table 1). Complementary size exclusion chromatography shows that IIa elutes as a single species over the practical loading concentration range of $10 \mu\text{M}$ to 1 mM (Fig. 2A,b); this species is identified as the dimeric IIa_2 observed by sedimentation equilibrium. Thus, the dimeric molecular mass of IIa_2 provides a benchmark for the size exclusion column and allows an estimate of the molecular masses of the remaining peptides. Figure 2A,a shows that the single- α -helical peptide p-Ia, which contains a blocked Cys thiol, elutes as a monomer or tetramer depending on the loading concentration. Hence, inter-chain disulphide loops are not needed to form the four-helix structure. Dimeric or trimeric assembly intermediates were not observed. Interestingly, the more hydrophilic peptide IId forms less stable dimers than IIa; monomers were observed at low loading concentrations (Fig. 2A,c). Addition of ferric haem to each of these peptides seems to stabilize the four- α -helical forms (Fig. 2A,d-f); an increase in the apparent molecular mass commensurate with the mass of haem was observed (Table 1). Addition of excess haem to IIa_2 at high concentrations produced a small population of higher-order aggregates (Fig. 2A,e), although the purified tetrahaem- IIa_2 was stable to further aggregation.

Circular dichroism (CD) spectroscopy shows that the structures of IIa_2 and IId_2 in aqueous solution (100 mM NaCl, 50 mM Tris pH 8) are highly α -helical (Table 1), close to the 90% expected from the sequence. In contrast, p-Ia₄ showed a concentration-dependent CD spectrum that fitted well to a

FIG. 2 Experimental characterization of the synthetic haem peptide. A, Size exclusion chromatograms of the peptides. A Pharmacia Superdex-75 (1×16 cm) size exclusion FPLC column, calibrated with a low molecular weight kit of globular proteins, was equilibrated with 50 mM Tris buffer, 100 mM NaCl, pH 8.0, and was eluted at a flow rate of 0.5 ml min^{-1} . Retention volumes for each peptide were determined by following the tryptophan absorbance at 280 nm. a–c Show elution profiles of peptides p-Ia, IIa and IIc, loaded at a concentration of $10 \mu\text{M}$. IIc₂ was observed (thinner trace; c) only at a higher loading concentration of 1 mM. d–f Show the effect of excess protoporphyrin IX on p-Ia, IIa and IIc. The molecular masses of these peptides and IIb and IIc (chromatograms not shown) so obtained are summarized in Table 1. B, Absorption spectra of the oxidized and reduced synthetic haem peptide. a Shows IIa₂ at $10 \mu\text{M}$ obtained in 50 mM Tris, pH 8.0, 100 mM NaCl on a Perkin-Elmer Lambda-2 spectrophotometer. One equivalent of protoporphyrin IX per IIa was added to fill the 10,10' sites and leave the 24,24' sites empty (see Table 1 for haem K_D values). A spectrum of free protoporphyrin IX in water is shown for comparison. In the oxidized form the γ -band maximum is 412.4 nm ($\epsilon 100 \text{ mM}^{-1} \text{ cm}^{-1}$) and the broad, α , β -band maximum is ~ 535 – 570 nm . The haems were reduced by addition of a minimal quantity of a freshly prepared dithionite solution yielding a reduced γ -band maximum at 425.2 nm ($\epsilon 150 \text{ mM}^{-1} \text{ cm}^{-1}$), β - and α -band maxima at 535 nm and 560 nm , respectively ($\epsilon 24 \text{ mM}^{-1} \text{ cm}^{-1}$). These maxima are nearly identical to those of *bis*-imidazole ferriprotoporphyrin IX in dimethylformamide solution (not shown). b Shows linear dichroic spectra of oriented Langmuir–Blodgett multilayer films of oxidized dihaem- (dashed line) and tetrahaem- (solid line) IIa₂. In linearly polarized light making a 20° incidence to the plane of the multilayer, absorption was greater when the plane of polarization was in the plane of the multilayer (upper dashed and solid traces). The dichroic ratios at 412 nm between the two polarizations indicate a haem orientation $\sim 25^\circ$ from the plane of the substrate for both the dihaem and tetrahaem peptide. The orientation is approximately parallel to the expected orientation of the helix axes in a compressed film. Pressure–area isotherms (data not shown) indicate that the peptide forms stable monolayers at the air–water interface. The peptide was transferred to quartz slides ($12 \times 25 \text{ mm}$) by repeated dipping through the air–water interface at surface pressures $\sim 20 \text{ mN m}^{-2}$. C, Contour plot of an expansion of a two-dimensional total correlation spectrum of the fully oxidized tetrahaem-IIa₂ protein. The bracketed region indicates those cross-peaks expected to arise primarily from α – β hydrogen resonance correlations. The spectrum was obtained at 500 MHz on a Bruker AMX-500 using a 0.75 mM sample



at 30°C in 5 mM potassium phosphate and 50 mM sodium chloride, pH 7.1. A standard pulse sequence⁴² was used, with an MLEV-17 mixing sequence of 19 ms. Each free induction decay consisted of 96 acquisitions of 512 complex points. A sweep width of 6,410 Hz was used in both the real time and incremented time domains. A data set of 600 incremented time domain values was acquired using time proportional phase incrementation⁴³ and processed with standard data reduction methods using the program Felix (Biosym).

tetramer–monomer equilibrium, with a dissociation constant (K_D) of $\sim 10 \text{ nM}^3$ and helical contents of 85% and 6% for the high- and low-concentration limits, respectively. The IIc₂ peptide shows similar behaviour (dimer–monomer equilibrium) in the presence of the chemical denaturant guanidinium chloride (Table 1).

Haem spectra

Addition of ferric haem to IIa₂ produced a spectrum (Fig. 2B,a and Table 1) typical of proteins containing *bis*-histidine-ligated ferric haem, with the characteristic Soret or γ -band near 412 nm and a broad α , β -band at 535 nm . Reduction to the ferrous form yielded the characteristic redshifted γ -band and prominent β -

and α -bands in the 530-nm and 560-nm regions, respectively. The ratio of the γ - and α -band extinction coefficients (6:1) is also typical of reduced *b*-type cytochromes. Moreover, the widths of the absorption bands contributing to the spectrum are consistent with values for the $\text{cyt } b_H$ of cytochrome *bc*₁, and suggest that the haems are in well-defined environments. For example, full-width at half-height of the α -band of ferrodhaem-IIb₂ is 11 nm , compared with 9 nm for the α -band of $\text{cyt } b_H$ (ref. 21). Haem addition to IIb₂, IIc₂ and IIc₂ yielded similar results (Table 1). Interestingly, the reaction of ferric haem with p-Ia₄ gave a nearly identical spectrum (Table 1), suggesting that the disulphide links in the II-type peptides are not essential for organizing the four-helix bundle to bind haem.

TABLE 1 Molecular masses and secondary structure of the synthetic peptides

Peptide	Calc.*	Molecular mass ($\times 10^{-3}$)				Circular dichroism ($-\theta_{222} \times 10^{-3}$ deg cm ² dmol ⁻¹) (% α -helix)	
		-Haem	+Haem		Sedimentary equilibrium‡	High conc. limit	Low conc. limit
		Chromatography†	Sedimentation equilibrium‡	Chromatography†			
p-la	3.846	4.2 $\pm$ 0.4	—	16 $\pm$ 2	—	27 $\pm$ 2§ (85 $\pm$ 10)	2 $\pm$ 2 (6 $\pm$ 6)
IIa	7.578	15 $\pm$ 2 15.156	14.9 $\pm$ 1.5¶	18 $\pm$ 2	18.7 $\pm$ 1.9¶	26 $\pm$ 2¶ (80 $\pm$ 10)	—
IIb	7.446	14 $\pm$ 2	—	17 $\pm$ 2	—	—	—
IIc	7.446	14 $\pm$ 2	—	16 $\pm$ 2	—	—	—
IId	7.668	8.3 $\pm$ 1 14 $\pm$ 2	14.8 $\pm$ 1.5☆	14 $\pm$ 2	20.9 $\pm$ 2.1☆	23 $\pm$ 2** (70 $\pm$ 10)	0.8 $\pm$ 0.8** (3 $\pm$ 3)

The molecular mass of iron protoporphyrin IX (haem) is 617. Circular dichroism spectra were obtained in 10 mM potassium buffer at pH 8.0 without added haem. 100% helicity is represented by a value for $-\theta_{222} \times 10^{-3}$ of 32 deg cm² dmol⁻¹ (ref. 38).

* Formula weight based on the sequence of the monomeric peptide.

† Apparent molecular masses were determined by size exclusion chromatography on a Pharmacia Superdex-75 column calibrated with the Pharmacia small molecular weight kit and the synthetic IIa₂ peptide. In cases where two values are given for a single peptide, the monomer and an aggregate (dimer IId₂ or tetramer p-la₄) were observed in either the same run or in independent runs at different loading concentrations.

‡ The calculated molecular masses determined by sedimentation equilibrium depend on the value of the partial specific volume, $\bar{v}$, used. Based on the sequence and published values, $\bar{v}$ was estimated to be 0.74 $\pm$ 0.025, giving a 10% uncertainty in the molecular masses.

§ Obtained at a total peptide concentration of 15 μ M.

|| Formula weight based on the sequence of the dimeric peptide.

¶ Average value obtained from two runs at loading concentrations of 2.5 μ M and 30 μ M. Within the cell the concentration covered the range ~1–60 μ M.

☆ Value obtained from one run at a loading concentration of 8 μ M. Within the cell the concentration range was ~2–20 μ M.

¶|| Obtained at total peptide concentration of 20 μ M. Addition of haem did not substantially affect the CD spectrum.

** Obtained in 2.5 M guanidinium chloride at peptide concentrations of 0.5–70 μ M.

Closer examination of the absorption spectra of IIb₂ (haems only at sites 10,10') and IIc₂ (haems only at sites 24,24') revealed oxidized and reduced spectra with different wavelength maxima. The γ - and α -band peak positions are listed in Table 1. For instance, the characteristic α -band maxima of the reduced minus oxidized forms are close to 558 nm for IIb₂ and 560 nm for IIc₂. Similar examination of IIa₂ with haems at both 10,10' and 24,24' sites reflects the presence of the two spectral forms measured individually in IIb₂ and IIc₂, suggesting the absence of significant electronic interactions between haems at the 10,10' and the 24,24' sites.

Haem affinities and molecular organization

Assay of the increase in absorption of the γ -band maximum on stepwise addition of ferric haem shows that, as expected, four haems bind per IIa₂ molecule and two haems per IIb₂ and IIc₂. Detailed titrations (not shown) under conditions where oligomer formation of IIb₂ and IIc₂ was minimized shows that one haem binds with a significantly higher affinity than the other, demon-

strating negative binding cooperativity. The K_D values (Table 2) are <0.05 μ M and 0.8 μ M for haem binding to the two 10,10' sites of IIb₂ and 0.8 μ M and 20 μ M for binding to the 24,24' sites of IIc₂, indicating that the free energy of haem–haem negative cooperativity is >1.63 and 1.88 kcal mol⁻¹, respectively. Similar analysis of IIa₂ shows that binding occurs over the same concentration range (individual K_D values not resolved), showing that there is little further negative cooperativity resulting from haem binding at the 10,10' and 24,24' sites on the same molecule.

Oriented monolayer and multilayer arrays of IIa₂ containing two haems (in the 10,10' high-affinity sites) and four haems (in both the 10,10' and 24,24' sites) have been deposited onto planar quartz substrates using conventional Langmuir–Blodgett (LB) techniques^{22,23}. Absorption spectra (Fig. 2B,b) of these multilayers display pronounced linear dichroism. The dichroic ratio of the Soret band indicates that all haems in the dihaem- and tetrahaem-containing IIa₂ peptides are nearly parallel (~25°) to the plane of the substrate. Compression of the film apparently orientates the α -helices nearly parallel to the substrate, with bis-

TABLE 2 Dissociation constants and spectroscopic and electrochemical properties of the synthetic multi-haem peptides

Peptide	Haem site	Haem dissociation constants (μ M)	Haem absorption maxima γ -band, α -band (nm)		Haem midpoint potentials (mV)*
			Oxidized	Reduced	
p-la ₄	10,10'	$\leq 1^\dagger$	412.7, 534.4	—	—
	24,24'	5 $\pm$ 4 [†]	—¶	—	—
IIa ₂	10,10'	0.2 $\pm$ 0.1 [†]	412.4, 534.7	425.2, 558.0	-130, -230
	24,24'	3 $\pm$ 2 [†]	—¶	427.5, 560.5	-80, -180
IIb ₂	10,10'	≤ 0.05 , 0.8 $\pm$ 0.10 [‡]	412.0, 533.8	424.5, 558.3	-100, -215
IIc ₂	24,24'	0.8 $\pm$ 0.3, 20 $\pm$ 10 [‡]	410.4 $\pm$ 532.5	427.2, 559.8	-75, -205
IId ₂	10,10'	≤ 0.005 , 0.35 $\pm$ 0.05 [‡]	414.0, 535.0	—	—
	24,24'	—§	—¶	—	—

* Redox titrations were conducted in 50 mM Tris, 100 mM NaCl, pH 8.0 at 25 °C (Fig. 3).

† Average K_D value for the two haems at the 10,10' site and the 24,24' site; no attempt was made to resolve individual binding constants.

‡ Two K_D values were resolved for the two haems at the 10,10' site and the 24,24' site.

§ The haem titration was conducted at such low peptide concentration (0.29 μ M) that only the two tight K_D values were measured; the 24,24' haem K_D values have not been determined.

|| Peak maxima were determined by adding 0.1 equivalent of haem per dimer so that no free haem was observed in the spectra.

¶ Difficult to determine accurately because of the presence of large free-haem absorption.

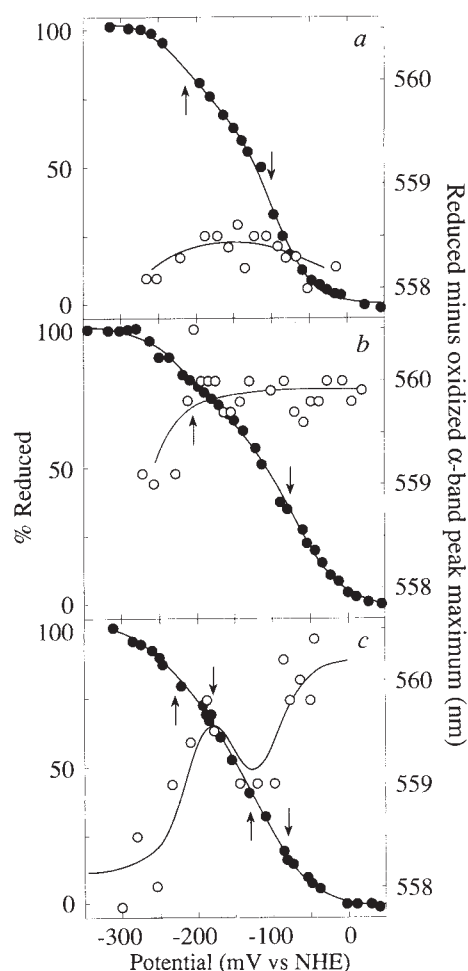


FIG. 3 Electrochemistry of haem peptides. Redox titrations of dihaem-IIb₂ (a), dihaem-IIc₂ (b) and tetrahaem-IIa₂ (c) are shown. Chemical titrations were done in a cuvette with platinum measuring and calomel reference electrodes²⁷. Ambient redox potential (measured against the standard hydrogen electrode; SHE) was adjusted by addition of small aliquots (<1 μ l) of sodium dithionite (reductant) or potassium ferricyanide (oxidant). Titrations were done in 50 mM Tris, pH 8.0, 100 mM NaCl. Mediation between the electrodes, the solution and the redox centres was facilitated with the following redox mediators: 10 μ M each *N*-methyl-1,4-naphthoquinone, 1,4-naphthoquinone, 1,4-benzoquinone; 15 μ M each 2-hydroxy-1,4-naphthoquinone, 1,4-naphthoquinone, 1,4-benzoquinone; 40 μ M 2,3,5,6-tetramethyl-*p*-phenylene diamine and duroquinone. After setting each potential, a spectrum was taken between 300 and 700 nm. An oxidized reference spectrum was subtracted from the spectrum obtained at each potential and the intensity of the α -band peaks was recorded. Absorbance (solid circles) was plotted against redox potential and the data were fitted to the Nernst equation with the indicated number of components. The n values for the data in a (IIb₂) and b (IIc₂) were best fitted to values of 1.0 and 0.75, respectively. The contributions of each of the two electrochemical states in a and b to the total α -band absorption are 35:65, respectively. Included in each data set is the wavelength maximum of the α -band (right ordinate, open circles) as a function of the redox potential. Spectral sampling was done by subtracting the spectrum taken at a higher potential from that taken immediately negative of it (typically obtained over a 10-mV potential range). c, A titration of the tetrahaem-IIa₂ showing the reduction of four discrete components and the concomitant alternation of wavelength maxima with potential. The absorption maximum dependence on potential was most simply modelled using four equally spaced redox midpoints alternating between haems with maxima at 560.5 and 558 nm. Because of the number of closely spaced electrochemical states no attempt was made to obtain n values.

histidine ligation forcing the haem groups parallel to the helices. This demonstrates that dihaem- and tetrahaem-IIa₂ are ordered structures and that, despite the negative cooperativity and the wide variation in the binding affinities of the four haem groups (>400-fold), they are oriented similarly within the frame of the four helices.

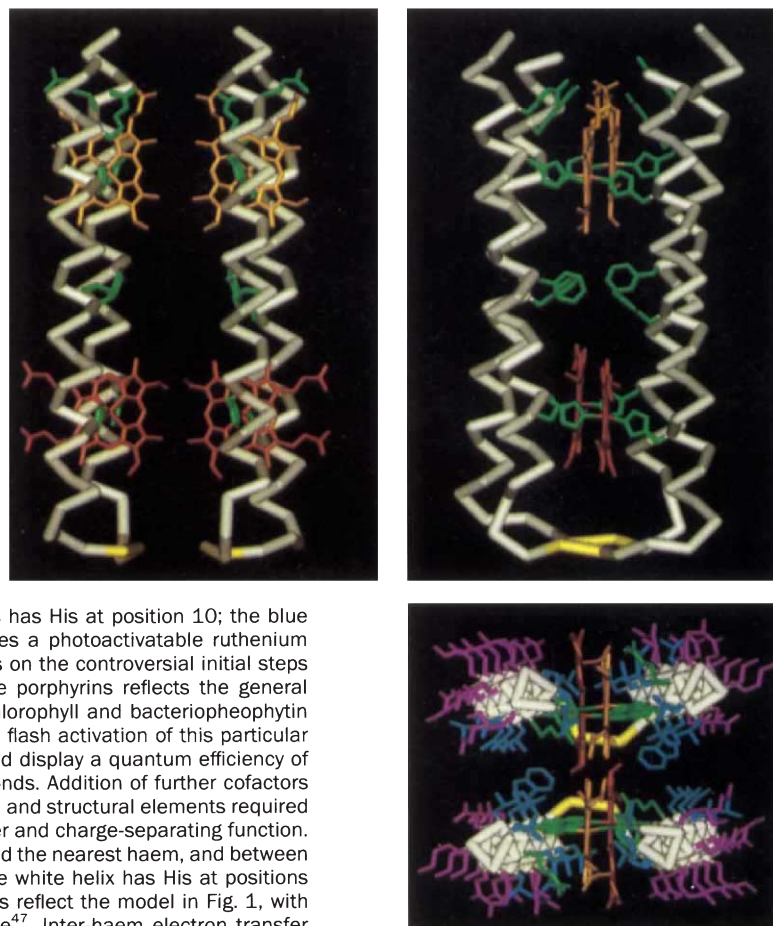
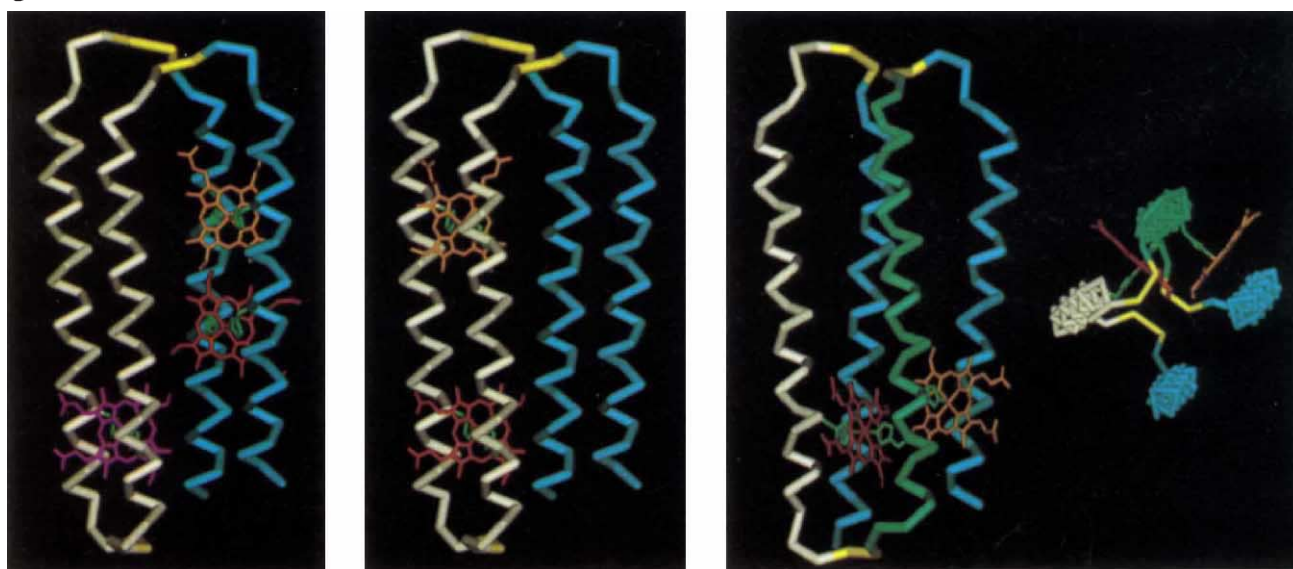
The electron paramagnetic resonance (EPR) spectrum of ferric tetrahaem-IIa₂ (not shown) is highly rhombic, with g -values (g_z , g_y , g_x) = 2.89, 2.24, 1.54, respectively), typical of *bis*-histidine-ligated low-spin (spin 1/2) haems²⁴, similar to that of *bis*-imidazole ferric haem. Furthermore, the spectrum of the tetrahaem-IIa₂ indicates that the ring planes of all the imidazole pairs ligating each haem are approximately parallel, in contrast to cytochrome *bc*₁ in which the imidazoles are predicted to be closer to perpendicular (see refs 24–26).

Figure 2C shows an expansion of a two-dimensional total correlation ¹H nuclear magnetic resonance (NMR) spectrum of the fully oxidized tetrahaem-IIa₂ obtained with a short isotropic mixing time. The corresponding two-dimensional double quantum-filtered correlation spectrum is essentially devoid of cross-peaks in the $H\alpha$ – $H\beta$ correlation region (not shown). This is attributed to the line broadening caused by the size of the complex (17.6 K) and the paramagnetic relaxation due to the presence of four spin-1/2 haems. The double quantum-filtered correlation spectrum is particularly sensitive to line broadening owing to the antiphase fine structure of cross-peaks. In contrast, the total correlation spectrum is less sensitive to line-width effects owing to its in-phase cross-peak fine structure. Under the conditions of large line widths and short isotropic mixing time, only directly J-coupled ¹H nuclei are expected to be correlated. The general appearance of the total correlation spectrum indicates that it arises from a population predominantly composed (>90%) of a single species. Although the presence of four spin-1/2 haem moieties in the protein assembly is expected to distort the expected chemical shift distribution somewhat, it is interesting that the region of the spectrum containing predominantly $H\alpha$ – $H\beta$ chemical shift correlations show about 60 intra-residue correlations. As the dimerized four-helix protein consists of 124 amino acids, this implies the presence of 2-fold symmetry in the complex. Note that ligation of the haem groups and the presence of disulphide bonds voids the potential for 4-fold symmetry in this protein. Figure 2C shows many other features that are the result of a unique structure. The detailed analysis of these data will be reported elsewhere.

Haem–haem electrochemical cooperativity

The electrochemical redox potentials of the bound haems in IIa₂, IIb₂ and IIc₂ were determined by monitoring the changes in the γ - and α -optical absorption bands as a function of redox potential²⁷. Figure 3a and Table 2 show that the haems at the 10,10' sites of the dihaem-IIb₂ assume two distinct electrochemical states, with midpoints at –215 mV and –100 mV. The two electrochemical states are coincident with the high- and low-affinity binding of ferrihaem to IIb₂ and hence may have the same origin. Although the two midpoints and the negative binding cooperativity may result from steric interactions pushing the haems into distinct local environments, an attractive hypothesis based on electrostatics provides a common basis for both effects. We propose that the positive charges of the two ferric haems electrostatically assist haem reduction, raising the midpoint (–100 mV)²⁸. Once one haem of the pair is reduced and the positive charge neutralized, the remaining haem adopts the lower midpoint of a non-interacting haem (–215 mV). Analogously, we propose that electrostatic repulsion between the two ferric haems leads to weak second haem binding. The 115-meV redox charge interaction free energy is equivalent to 2.5 kcal mol^{–1}, which is comparable to the >1.63 kcal mol^{–1} difference in binding free energies of the two haems to the 10,10' sites of IIb₂. Consistent with this hypothesis, redox titration of a similar four-helix structure containing only one *bis*-histidyl

FIG. 4 *a*, Model of the favoured structure of the tetrahaem-*Ila*₂ protein, showing front, side and top views, derived from UV–visual absorption, electrochemical, EPR and NMR data. The C α -backbone trace is shown in white, the Cys1–Cys1' disulphide bonds in yellow, the haems at positions 10,10' in red, the haems at positions 24,24' in orange relevant side chains (His 10, Phe 17, His 24 and Arg 27) in green, and in the top view the hydrophobic residues (Trp, Leu) are shown in blue and the charged residues (Glu, Lys) in purple. The helix–helix separations (~ 45 Å) and relative angles (0 – 10°) in the model are defined by the constraints, derived from the EPR data, placed on the histidine dihedral angles χ_1 and χ_2 . Phe, Trp and Leu side-chain conformations are derived by minimizing protein–protein and protein–haem interactions. *b*, Maquette designs for the photosynthetic reaction centre, cytochrome *bc*₁ and cytochrome *c* oxidase. These structures are nearly identical to that of *a*, with a simple modification of the primary sequence. A 31-amino-acid type-I peptide with two histidines is linked as before through a disulphide bond at the N termini to another 31-amino-acid peptide with either one histidine (reaction centre) or no histidines (cytochrome *bc*₁). The resultant 62-amino-acid peptide is linked through a disulphide bond at the carboxyl termini to create a 4-helix bundle. The cytochrome *c* oxidase maquette requires a similar linkage of three different type-I peptides. These hetero links are easily made using established synthetic methods⁴⁴. Reaction centre maquette: The white helix tracing the peptide C α -atoms has His at position 10; the blue helix has His at positions 17 and 24. This maquette comprises a photoactivatable ruthenium porphyrin (purple) and two haems (red and orange) and focuses on the controversial initial steps of photosynthesis. The overall geometrical arrangement of the porphyrins reflects the general positioning of the bacteriochlorophyll dimer and the bacteriochlorophyll and bacteriopheophytin of the bacterial reaction centre¹. We calculate^{37,46} that a single flash activation of this particular maquette, using haems with potentials as described here, would display a quantum efficiency of $>90\%$ and a lifetime of the charge-separated state of milliseconds. Addition of further cofactors would extend function and reveal the minimal physical, chemical and structural elements required for bioenergetic oxidation–reduction, long-range electron transfer and charge-separating function. The edge-to-edge distances between the ruthenium porphyrin and the nearest haem, and between the haems are 8 Å and 5 Å, respectively. Cyt *bc*₁ maquette: The white helix has His at positions 10, 24, while the blue helix has no His. Distances and locations reflect the model in Fig. 1, with accompanying helices of part of a cytochrome *b* helical bundle⁴⁷. Inter-haem electron transfer can be activated by attaching photoactivatable groups such as Ru(bpy)₂(im) (refs 48, 49) at strategic points on the peptide, or by applying electric fields to oriented monolayers^{22,23}. This brings new opportunities for examining this central reaction of cytochrome *bc*₁, which has been kinetically inaccessible in the native protein and complicated by the presence of other cofactors. The haem edge-to-edge distance is 16 Å. Cyt *c* oxidase maquette: The white helix has His at positions 10, the green helix has His at positions 10 and 13, and the blue helix has no His. The design is based on working models^{4,6} and will facilitate the description of the interactions between two haems attached to the same helix, the low-spin six-coordinate cytochrome *a*, and the high-spin five-coordinate cytochrome *a*₃. The top view reveals the similarities to the haem spatial positioning and coordination environments in these working models. The Fe–Fe distance is 10.4 Å and the edge-to-edge distance is 5.2 Å. Preliminary work has prepared high-spin haems in similar four-helix proteins (R.S.F., unpublished results, and ref. 45).

a*b*Reaction centre
maquetteCyt *bc*₁
maquetteCyt *c* oxidase
maquette

ligated haem shows a single midpoint close to the lower value (-220 mV, ref. 45). Furthermore, redox titrations of a *bis*-histidine-ligated single-haem octapeptide displays a midpoint near -215 mV at pH 7 (refs 29, 30).

Figure 3a also shows that the redox potential dependency of each redox transition in the dihaem-IIb₂ fits well to a theoretical Nernstian n value of 1.0 indicative of a well-defined two-state process. This suggests that the haem groups exist in a homogeneous electrochemical environment, consistent with the high affinities and well-defined spectra of both haems binding to similar 10,10' sites. In addition, the position of the reduced haem α -band maximum, sampled incrementally during redox titration, stays within a range of 0.5 nm (Fig. 3a); thus, the effects of interaction between haems in the 10,10' sites on the spectroscopic properties are relatively slight.

Figure 3b shows that the haems at the 24,24' sites of dihaem-IIc₂ exhibit analogous electrochemical behaviour to that found at the 10,10' sites of dihaem-IIb₂. The midpoints of the haems of dihaem-IIc₂ are slightly but significantly raised (-205 mV and -75 mV), as expected from the placement of positively charged Arg 27,27' on the outer surface of the peptide near the 24,24' haem site. A similar effect of arginine on the midpoint of cytochrome *c* (raised by ~ 30 mV) has been reported¹⁸. The difference between the two midpoints indicates a charge-interaction free energy of 130 meV or 2.9 kcal mol⁻¹. The corresponding negative cooperative binding interaction (Table 2) is equivalent to 1.9 kcal mol⁻¹, which suggests that most of the effects on binding can be ascribed to charge interactions.

The n values of the redox titration for the two haems in dihaem-IIc₂ are both smaller than for dihaem-IIb₂, ~ 0.75 , consistent with a more heterogeneous haem-binding interaction at the low-affinity 24,24' site, and indicative of a normal distribution of electrochemical values with a standard deviation of 25 mV. Finally, the spectral sampling in dihaem-IIc₂ indicates a 0.8-nm shift to shorter wavelengths after the first haem is reduced. This suggests that there are redox-coupled spectral interactions between the haems at a greater level than with dihaem-IIb₂.

Figure 3c shows an identical redox potentiometric examination of the four haems in tetrahaem-IIa₂ (that is, two haems in the 10,10' and two in the 24,24' sites). A combined analysis of the extent of reduction with spectral sampling resolves four electrochemical states with the absorption band maxima and midpoint potential values given in Table 2. This titration is not quite a superposition of the simpler dihaem-IIb₂ and IIc₂ titrations; further perturbation of the midpoints occurs, presumably because of the proximity of the two pairs of haems. Reduction of the four haems proceeds as follows: first, a 560-nm haem (24,24') is reduced at -80 mV; second, a 558-nm haem (10,10') at -130 mV; third, another 560-nm haem (24,24') at -180 mV; and then the last, the other 558-nm haem (10,10'), with a midpoint at -230 mV close to that characteristic of a non-interacting haem.

Structural considerations

Electrostatic interactions between haems suggested by binding and midpoint values have implications for the overall structure of the haem peptides. Our analysis indicates that the tetrahaem-IIa₂ and dihaem-IIb₂ and -IIc₂ consist of two identical monomer subunits conforming to the four-helix bundle motif, a structure that readily accommodates haems oriented with nearly parallel molecular planes and that displays a 2-fold symmetry. The two most likely 2-fold symmetric structures depend on the forces governing the association. Figure 4a shows a possible structure of tetrahaem-IIa₂ with α -helix macrodipoles in an all-parallel configuration and hence the disulphide links at the same end of the structure. The other main alternative has the macrodipoles of the helices of the two monomers in an antiparallel configuration, with the disulphide links at opposite ends.

We favour the parallel model because it brings the edges of

the apparently interacting dihaem-IIb₂ and -IIc₂ into van der Waals contact with an Fe-to-Fe distance of 13 Å. At this distance, charge interactions through the interior of a water-soluble protein can be reasonably expected to be in the range of 100 mV³¹. In contrast, the antiparallel model has a haem Fe-to-Fe distance of 26 Å along the axis of the molecule, a distance over which the high aqueous dielectric value will be expected to dominate the low internal dielectric value and diminish the charge interactions to <5 mV. Tetrahaem-IIa₂, with two pairs of haems in close contact, has a structurally revealing sequence of haem reduction. The parallel model geometry predicts that the first reduction would be predominantly one of the haems of the 24,24' pair assisted by the additional positive charges of Arg at 27 and 27'. The interacting positive charges remaining on the 10,10' pair would lead to the reduction of one of these haems. With one of each pair reduced, the Arg effect suggests that the remaining ferric haem in a 24,24' site will be the haem predominantly reduced next, followed finally by the ferric haem at the 10,10' site. This is the sequence of reduction shown in Fig. 3c. In contrast, the antiparallel model would predict that the two 24,24' ferric haems would be reduced first, followed by the two 10,10' ferric haems at the lower potential, leading to a spectral signature sharply different from that found in Fig. 3c.

Initially we were concerned that the protein might suffer from excessive conformational heterogeneity and dynamic behaviour, leading to a series of nondescript properties. However, despite the simplicity of their construction, the di- and tetrahaem protein assemblies show a remarkable degree of structural, spectroscopic and electrochemical definition. Indeed, the synthetic haem proteins display a sophisticated set of properties of the kind long proposed to be of functional importance in the more complex natural redox enzymes and energy conversion proteins. Particularly noteworthy are the negatively cooperative haem-haem redox interactions first proposed in cytochrome oxidase^{32,33} and considered vital for its proton pumping⁵, although not, we believe identified with a simple electrostatic source. This phenomenon has also been proposed in dihaem cytochrome *c*₄ (ref. 28) and tetrahaem cytochrome *c*₃ (refs 34–36).

Maquette design

The synthesis of this prototypical family of multi-haem proteins shows how we can now freely place redox centres in proteins, unconstrained by many of the complexities of biological systems. Figure 4b shows maquette designs for three central charge-separating redox enzymes: photosynthetic reaction centre, with an added photoactivatable ruthenium porphyrin; cytochrome *bc*₁, with a pair of six-coordinate haems; and cytochrome oxidase, with one member of the haem pair having the sixth coordination position free for ligand binding. Each maquette is based on two different di- α -helical monomers (three different monomers for the cytochrome oxidase maquette) linked to form four-helix bundles using established technology¹⁹, with redox centre placement appropriate for intra-protein electron transfer³⁷. The simplicity of these maquettes provides an unusually tractable experimental environment for discovering the minimal requirements for redox protein function. □

Received 18 October 1993; accepted 27 January 1994.

1. Michel, H., Deisenhofer, J. & Epp, O. *EMBO J.* **5**, 2445–2451 (1986).
2. Robertson, D. E. & Dutton, P. L. *Biochim. biophys. Acta* **935**, 273–291 (1988).
3. Crofts, A. R., Robinson, H., Andrews, K., van Doren, S. & Berry, E. in *Cytochrome Systems: Molecular Biology and Bioenergetics* (eds Papa, S., Chance, B. & Ernster, L.) 617–624 (Plenum, New York, 1987).
4. Hosler, J. P. et al. *J. Bioenerg. Biomembranes* **25**, 121–136 (1993).
5. Babcock, G. T. & Wikström, M. *Nature* **356**, 301–309 (1992).
6. Brown, S., Moody, A. J., Mitchell, R. & Rich, P. R. *FEBS Lett.* **316**, 216–233 (1993).
7. Mitchell, P. *Nature* **191**, 144–148 (1961).
8. Mitchell, P. *J. theor. Biol.* **62**, 327–367 (1976).
9. Handel, T. M., Williams, S. A. & DeGrado, W. F. *Science* **261**, 879–885 (1993).
10. Betz, S. F., Raleigh, D. P. & DeGrado, W. F. *Curr. Opin. struct. Biol.* **3**, 601–610 (1993).
11. Glaser, E. G. & Crofts, A. R. *Biochim. biophys. Acta* **766**, 322–333 (1984).
12. West, I. C., Mitchell, P. & Rich, P. R. *Biochim. biophys. Acta* **933**, 35–41 (1988).
13. Saraste, M. *FEBS Lett.* **166**, 367–372 (1984).
14. Howell, N. & Gilbert, K. in *Cytochrome Systems: Molecular Biology and Bioenergetics* (eds Papa, S., Chance, B. & Ernster, L.) 79–86 (Plenum, New York, 1987).

15. Yun, C. H., Crofts, A. R. & Gennis, R. B. *Biochemistry* **30**, 6747–6754 (1991).
16. Moore G. *FEBS Lett.* **161**, 171–175 (1983).
17. Cramer, W. A., Black, M. T., Widger, W. R. & Girvin, M. E. *Structure and Function of Photosynthetic Cytochrome bc1 and b6f Complexes* (Elsevier, Amsterdam, 1987).
18. Davies, A. M. et al. *Biochemistry* **32**, 5431–5435 (1993).
19. DeGrado, W. F., Wasserman, Z. R. & Lear, J. D. *Science* **243**, 622–628 (1989).
20. DeGrado, W. F., Raleigh, D. P. & Handel, T. *Curr. Opin. struct. Biol.* **1**, 984–993 (1991).
21. Howell, N. & Robertson, D. E. *Biochemistry* **32**, 1310–1317 (1993).
22. Alegria, G. & Dutton, P. L. *Biochim. biophys. Acta* **1057**, 239–257 (1991).
23. Alegria, G. & Dutton, P. L. *Biochim. biophys. Acta* **1057**, 258–272 (1991).
24. Palmer, G. *Biochem. Soc. Trans.* **13**, 548–560 (1985).
25. Salerno, J. C. *J. biol. Chem.* **259**, 2331–2336 (1984).
26. Walker, F. A., Huynh, B. H., Scheidt, W. R. & Osvath, S. R. *J. Am. chem. Soc.* **108**, 5288–5297 (1986).
27. Dutton, P. L. *Meth. Enzym.* **54**, 411–435 (1978).
28. Leitch, F. A., Brown, K. R. & Pettigrew, G. W. *Biochim. biophys. Acta* **808**, 213–218 (1985).
29. Harbury, H. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **54**, 1658–1664 (1957).
30. Churg, A. K. & Warshel, A. *Biochemistry* **25**, 1675–1681 (1986).
31. Gunner, M. R. & Honig, B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9151–9155 (1991).
32. Nicholls, P. & Petersen, L. C. *Biochim. biophys. Acta* **357**, 462–467 (1974).
33. Wikstrom, M. K. F., Harmon, H. J., Ingledew, W. J. & Chance, B. *FEBS Lett.* **65**, 259–277 (1976).
34. Moura, J. J. G. et al. *Eur. J. Biochem.* **127**, 151–155 (1982).
35. Santos, H., Moura, J. J. G., Moura, I., LeGall, J. & Xavier, A. V. *Eur. J. Biochem.* **141**, 283–296 (1984).
36. Fan, K., Akutsu, H., Kyogoku, Y. & Niki, K. *Biochemistry* **29**, 2257–2263 (1990).
37. Moser, C. C., Keske, J. M., Warncke, K., Farid, R. S. & Dutton, P. L. *Nature* **355**, 796–802 (1992).
38. Pace, C. N., Shirley, B. A. & Thompson, J. A. in *Protein Structure: A Practical Approach* (ed. Creighton, T. E.) 311–330 (IRL, Oxford, 1989).
39. Dutton, P. L. & Jackson, J. B. *Eur. J. Biochem.* **30**, 495–510 (1972).
40. Orme-Johnson, N. R., Hansen, R. E. & Beinert, H. *J. biol. Chem.* **249**, 1928–1939 (1974).
41. Fields, G. B. & Noble, R. L. *Int. J. Peptide Protein Res.* **33**, 1–53 (1989).
42. Bax, A. & Davis, D. G. *J. magn. Reson.* **65**, 355–360 (1985).
43. Marion, D. & Wuthrich, K. *Biochem. biophys. Res. Commun.* **113**, 967–974 (1983).
44. O'Shea, E. K., Rutkowski, R., Stafford, W. F. & Kim, P. S. *Science* **245**, 646–648 (1989).
45. Choma, C. T. et al. *J. Am. Chem. Soc.* **116**, 856–865 (1994).
46. Moser, C. C. & Dutton, P. L. *Biochim. biophys. Acta* **1101**, 171–176 (1992).
47. Crofts, A., Hacker, B., Barquera, B., Yun, C.-H. & Gennis, R. *Biochim. biophys. Acta* **1101**, 162–165 (1992).
48. Willie, A., Stayton, P. S., Sligar, S. G., Durham, B. & Millett, F. *Biochemistry* **31**, 7237–7242 (1992).
49. Wuttke, D. S., Bjerrum, M. J., Winkler, J. R. & Gray, H. B. *Science* **256**, 1007–1009 (1992).

ACKNOWLEDGEMENTS. Funded by grants from the US PHS (P.L.D. and A.J.W.).

LETTERS TO NATURE

Discovery of an X-ray source coincident with the soft γ -ray repeater 0525 – 66

R. E. Rothschild*, S. R. Kulkarni† & R. E. Lingfelter*

* Center for Astrophysics and Space Sciences 0111, University of California, San Diego, La Jolla, California 92093–0111, USA

† Division of Physics, Mathematics and Astronomy 105–24, California Institute of Technology, Pasadena, California 91125, USA

ALTHOUGH γ -ray bursts (GRBs) have been known for more than 20 years, no source has ever been identified in its quiescent state, which might provide clues to its nature. On the other hand, two of the three known soft γ -ray repeaters (SGRs), which emit intermittent bursts of soft γ -rays, seem to be associated with supernova remnants^{1,2}, and the recent identification of X-rays from one of these, SGR1806 – 20, supports the suggestion that a pulsar inside the remnant is the source of the γ -rays^{3–5}. Here we report X-ray observations of SGR0525 – 66, which has been associated previously with the supernova remnant N49 (ref. 1). We identify point-like emission from a source coincident with SGR0525 – 66, which suggests that it too is a pulsar. The pulsar seems to be only about 5,000 years old and has a high transverse velocity of about $1,200 \text{ km s}^{-1}$, and we predict that the plerion (the region of radio synchrotron emission surrounding the pulsar) will be between 0.1 and 0.3 arcsec across. A high birth velocity has been estimated for the pulsar associated with SGR1806 – 20 also⁴, and this characteristic may be related to the reason why only a very few pulsars become SGRs.

The smallest γ -ray-burst error box is that for the 5 March 1979 event from the repeating burst source SGR0525 – 66 (ref. 6). Until recently this 0.09 arcmin^2 region was unique in that it overlapped the 2 arcmin^2 supernova remnant N49 in the Large Magellanic Cloud¹. Murakami *et al.* report⁵ that SGR1806 – 20 is also coincident with a supernova remnant, G10.0 – 0.3. Previous observations of N49 have reported a region of excess emis-

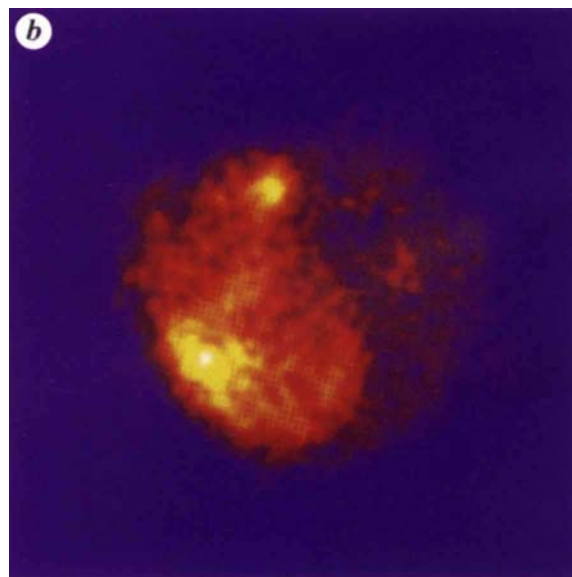
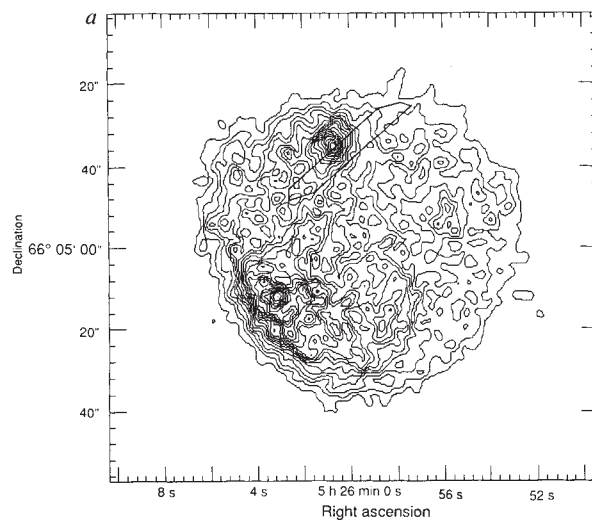


FIG. 1 a, Rosat High Resolution Imager smoothed contour plot of the supernova remnant N49 in the Large Magellanic Cloud. The point source at RA (J2000) 5 h 26 min 0.7 s and dec (J2000) –66° 04' 35" is coincident with the error box (six-sided polygon) for SGR0525 – 66 which is the site of the 5 March 1979 γ -ray burst. Twenty, equally

spaced, contour intervals are displayed, ranging from 0.72 to $14.60 \text{ smoothed counts arcsec}^{-1}$. b, False-colour image of N49 from which the contours of a are derived. This represents $\sim 20,000 \text{ s}$ of livetime to the source.

sion, possibly point-like, within the nebula and consistent with the γ -ray-burst error box⁷. The satellite observations from Rosat of N49 over 3.5 days (beginning 17 March 1992) with the High Resolution Imager (HRI) yielded just over 20,000 s of livetime on the source. The resultant image contains $1.67 \pm 0.014 \times 10^4$ net counts in the 0.1–2.4 keV band for an average counting rate from N49 of $8.04 \pm 0.07 \times 10^{-1}$ counts s^{-1} . The image was smoothed using the PROS/IRAF software package, resulting in the contours and false-colour image shown in Fig. 1a and b respectively. A point-like source of excess emission is clearly seen in the northern portion of the remnant, as well as a broad, structured feature to the southeast. This morphology is consistent with that seen by HRI on the Einstein satellite 13 yr earlier⁸. The Interplanetary Network γ -ray error box for the intense 5 March 1979 γ -ray burst⁶, shown in Fig. 1a (six-sided polygon), is consistent with the northern hotspot in N49. The emission from this hotspot is centered at right ascension (RA) (J2000) 5 h 26 min 0.7 s and declination (dec.) (J2000) $-66^\circ 04' 35''$ with an uncertainty of $\sim 10''$ in radius. The excess counting rate from the hotspot is $1.51 \pm 0.13 \times 10^{-2}$ counts s^{-1} above that from an annulus of $5''$ inner radius and $10''$ outer radius centred on it. This represents $\sim 2\%$ of the remnant's 0.1–2.4 keV emission.

A radial profile centred on the hotspot was generated out to $10''$ with background selected from the next $5''$ in the radial direction. The resulting profile is shown in Fig. 2, along with a fit to a point-like radial gaussian response function plus a constant term representing any unsubtracted nebular emission. The fitted width of the response, $\sigma = 3.19'' \pm 1.57''$ is consistent with the combined point spread function of the Rosat telescope and the HRI of $2.12''$, and, thus, the spatial distribution of counts in the hotspot is indeed consistent with a point source of emission.

A spectral form is necessary to estimate the flux from the point source. Because no spectral information is available from the Rosat HRI, we assume a neutral hydrogen column density of 5×10^{20} atoms cm^{-2} along the line of sight⁹, and we estimate¹⁰ the energy-to-counts conversion factor (ECF) to be 7.5×10^{-2} counts $s^{-1}/(10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1})$ based upon a Crab-like pulsar with energy index of 1, and to be 6×10^{-2} counts $s^{-1}/(10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1})$ for a black body with a temperature of $\sim 2 \times 10^6$ K appropriate for a 5×10^3 -yr-old neutron star¹¹. The Crab ECF results in a flux estimate of $\sim 2.0 \times 10^{-12}$ ergs $cm^{-2} \text{ s}^{-1}$, whereas that for the black body is $\sim 2.5 \times 10^{-12}$ ergs $cm^{-2} \text{ s}^{-1}$. At the distance of the Large Magellanic Cloud (55 kpc)¹², the luminosity is $\sim 7 \times 10^{35}$ erg s^{-1} for the power law and $\sim 9 \times 10^{35}$ erg s^{-1} for the black body.

We have looked carefully at the optical emission in the vicinity of the point source, and although there is [O II], H α , and 6,100-Å continuum emission in this region, the emission is faint and diffuse with no bright object coincident with the X-ray hotspot. A comparison with [Fe XIV] 5,303-Å observations¹³ yields a similar conclusion. The brightest X-ray emission regions to the southeast, on the other hand, correlate quite well with the brightest [Fe XIV] regions. The absence of such a strong correlation near the point source is, perhaps, the best observational evidence for a non-nebular origin for the majority of the point source X-ray flux. Using star 3 of a previous optical investigation of the SGR0525–66 error box¹⁴ for comparison, we estimate an upper limit to the brightness of an optical point source to be $m_V \sim 20$ mag.

If we assume that the point-like source coincident with the N49 nebula is indeed a neutron star born in the supernova event, as also appears to be the case in SGR1806–20, its $\sim 25''$ offset from the centre of the remnant implies a transverse velocity of $v_p \sim 1,200$ km s^{-1} , given the age $t = 5.4 \times 10^3$ year (ref. 15). Only 3% of the pulsar population have velocities exceeding 500 km s^{-1} , and the highest measured¹⁶ velocity is 10^3 km s^{-1} .

Calculations for the cooling of neutron stars¹¹ predict that after 5×10^3 yr the temperature will be $1\text{--}3 \times 10^6$ K and the bolometric luminosity will be in the range $2\text{--}50 \times 10^{33}$ erg s^{-1} . The present observations imply an X-ray luminosity (0.1–2.4 keV)

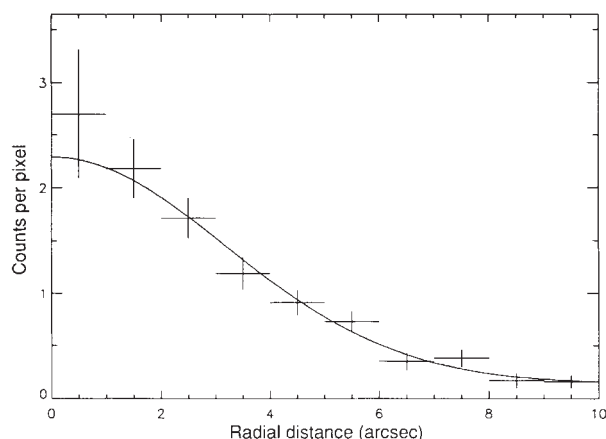


FIG. 2 Radial profile (data points) of the background-subtracted hotspot emission in N49 coincident with the γ -ray-burst error box. The background used was taken immediately surrounding the hotspot. The best-fit value of the width of the radial gaussian $\sigma = 3.19'' \pm 1.57''$ ($\chi^2_\nu = 0.71$ for $\nu = 7$ degrees of freedom) is consistent with the point source response value of $2.12''$. This best fit to the profile is given as the solid line.

of the order of 10^{36} erg s^{-1} . As the 0.1–2.4 keV flux for this temperature black body represents essentially the whole bolometric luminosity, a source of heating, such as γ -ray bursts or subsequent accretion, is required to account for the higher luminosity for the case of a black-body spectrum.

The existence of a non-thermal nebula, G10.0–0.3, around SGR1806–20 (ref. 4) indicates that soft γ -ray repeaters are also sources of high-energy particles and magnetic fields. Unlike most supernova remnants, G10.0–0.3 is centre filled and not a shell. The surface brightness increases towards the centre where quiescent X-ray emission has been detected by the ASCA satellite⁵. Two classes of neutron star systems have similar unusual radio nebulae: radio pulsars, such as the Crab nebula, and a rare type of luminous accreting X-ray binaries, (such as SS433 embedded in the nebula W50, and Cir X–1 (ref. 17)).

Our observations of SGR0525–66 can be interpreted in the context of these two possibilities for SGR1806–20. The persistent X-ray luminosity of the two are comparable. Because the line-of-sight extinction to the Large Magellanic Cloud is very low, unlike that to SGR1806–20, constraints on the model can be particularly strong. The absence of a bright optical star, as well as the persistent flux observed from SGR0525–66 being several orders of magnitude smaller than that of SS433 or Cir X–1, does not favour the model of an X-ray binary within a nebula. In the pulsar model, the X-ray point source can be interpreted as a compact synchrotron nebula, or a so-called plerion. A number of such supernova remnants with a plerion embedded within the shell exist in our Galaxy¹⁸. Assuming, as in other plerions, an efficiency $\eta \approx 10^{-2} \eta_{-2}$, where η_{-2} is the efficiency in units of 10^{-2} , for the conversion of spin-down luminosity ($\dot{E}$) into X-rays, we can infer $\dot{E} \propto B_{12}^2/P^4 \approx 7 \times 10^{37}$ erg s^{-1} , where B_{12} is the surface dipole field strength in units of 10^{12} G and P is the rotational period in seconds. Here we assume the simple magnetic dipole spin-down model for pulsars and $\eta_{-2} \approx 1\text{--}3$ for the known plerions¹⁹. The same model yields $P^2/B_{12}^2 = (t/17 \times 10^6 \text{ yr})$. From these two relations and our assumed $\dot{E}$ and t , we find $B_{12} = 3\eta_{-2}^{1/2}$ and $P = 45\eta_{-2}^{1/2}$ ms.

An 8.0-s periodicity was found²⁰ in the emission after the peak in the 5 March 1979 burst, and a 23-ms quasi-periodic oscillation was reported²¹ during the peak of that burst. An 8.0-s rotation period would require an efficiency $\eta_{-2} \gg 1$; however, the 23-ms period, if confirmed, would be quite consistent with the proposed model, giving $\eta_{-2} = 0.26$ and $B_{12} = 1.5$.

Our hypothesis is testable. The radius of the plerion (r) is determined by the balance between the ambient pressure and the momentum flux of the pulsar's relativistic wind, $\dot{E}/4\pi r^2 c$, where c is the velocity of light. The ambient pressure is taken to be the larger of the two possibilities: the mean pressure in the supernova remnant ($p \approx n_0 m_H v_{sh}^2$), or the ram pressure due to the motion of the pulsar close to the edge of the supernova remnant ($p \sim 4n_0 m_H v_p^2$). Here $v_{sh} = 700 \text{ km s}^{-1}$ is the blast wave speed, $n_0 = 1 \text{ atom cm}^{-3}$, the density of the medium into which the supernova remnant is expanding¹⁵ and m_H is the mass of a hydrogen atom. We predict the plerion to have a size between 0.1 arcsec (if ram pressure dominates) and 0.3 arcsec (if the mean pressure in the supernova remnant dominates). In the former (likely) case, the plerion should be distinctly cometary whereas for the latter a spherical plerion is expected. It appears to us that sensitive high-resolution radio imaging of this region

would be a very worthwhile undertaking. Similarly, radio and X-ray spectra of the source would be valuable, as the characteristic power-law signature should be present at some level.

Although the high transverse velocity for this pulsar is near the extreme of observed values, the magnetic field and spin period parameters are not very different from those deduced for the pulsar population as a whole, and we have not had to invoke high field strengths²². If, indeed, SGR0525–66 is a typical young pulsar, then what property makes it special to undergo and repeat soft γ -ray bursting? The spatial offset of both SGR0525–66 and SGR1806–20 from the centres of their respective supernova remnants has been noted⁴ and this may imply an unusual manner for their birth. The precise connection between soft γ -ray repeating and offsets (high velocities?) remains to be identified. $\square$

Received 10 November 1993; accepted 10 February 1994.

1. Helfand, D. J. & Long, K. S. *Nature* **282**, 589–591 (1979).
2. Kulkarni, S. R. & Frail, D. A. *Nature* **365**, 33–35 (1993).
3. Cooke, B. A. *Nature* **366**, 413–414 (1993).
4. Kulkarni, S. R., Frail, D. A., Kassim, N. E., Murakami, T. & Vasisht, G. *Nature* **366**, 129–131 (1994).
5. Murakami, T. et al. *Nature* **366**, 127–129 (1994).
6. Cline, T. L. et al. *Astrophys. J.* **255**, L45–L48 (1982).
7. Rothschild, R. E., Lingefelter, R. G., Seward, F. D. & Vancura, O. in *Compton Gamma-Ray Observatory AIP Conf. Proc.* Vol. 280 (eds Friedlander, M., Geherels, N. & Macomb, D.) 808–812 (Am. Inst. Phys. Press, New York, 1993).
8. Mathewson, D. S. et al. *Astrophys. J. Suppl. Ser.* **51**, 345–355 (1983).
9. Pizzichini, G. et al. *Astrophys. J.* **301**, 641–649 (1986).
10. ROSAT Mission Description, Appendix F, Report No. NRA 91-OSSA-3 (Max-Planck-Institut Phys. Astrophys., Garching, 1991).
11. Nomoto, K. & Tsuruta, S. *Astrophys. J.* **250**, L19–L23 (1981).
12. Capaccioli, M., Della Valle, M., D'Onofrio, M. & Rosino, L. *Astrophys. J.* **360**, 63–67 (1990).

13. Dopita, M. A. & Mathewson, D. S. *Astrophys. J.* **231**, L147–L150 (1979).
14. Fishman, G. J., Duthie, J. G. & Dufour, R. J. *Astrophys. Space Sci.* **75**, 135–143 (1981).
15. Vancura, O., Blair, W. P., Long, K. S. & Raymond, J. C. *Astrophys. J.* **394**, 158–173 (1992).
16. Harrison, P. A., Lyne, A. G. & Anderson, B. *Mon. Not. R. astr. Soc.* **261**, 113–124 (1993).
17. Stewart, R. T., Caswell, J. L., Haynes, R. F. & Nelson, G. J. *Mon. Not. R. astr. Soc.* **261**, 593–598 (1993).
18. Weiler, K. W. & Sramek, R. A. *Rev. Astr. Astrophys.* **26**, 295–341 (1988).
19. Seward, F. D. & Wang, Z.-R. *Astrophys. J.* **332**, 199–205 (1988).
20. Mazets, E. P. et al. *Nature* **282**, 587–589 (1979).
21. Barat, C. et al. *Astr. Astrophys.* **126**, 400–402 (1983).
22. Duncan, R. C. & Thompson, C. *Astrophys. J.* **392**, L9–L13 (1992).

ACKNOWLEDGEMENTS. We thank R. Petre, M. Corcoran, and the Rosat Data Center duty scientists for assisting with the understanding of the data and its analysis, and O. Vancura for providing the optical images of N49. R.E.R. and R.E.L. are supported by NASA. S.R.K. is supported by US NSF, NASA and the Packard foundation.

Possible role of massive black holes in the generation of galactic magnetic fields

Sandip K. Chakrabarti^{*†‡}, R. Rosner[†] & S. I. Vainshtein[‡]

^{*} Tata Institute of Fundamental Research, Bombay 400005, India

[†] International Centre for Theoretical Physics, 11 Strada Costiera, 34100 Trieste, Italy

[‡] Department of Astronomy & Astrophysics, University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637, USA

THE origin of galactic magnetic fields has been a long-standing puzzle. Models based on standard dynamo theory^{1–4} encounter several problems, the most fundamental of which is that, in order to explain the strengths of observed large-scale magnetic fields^{5–7}, the fluctuating magnetic fields in galaxies must be unreasonably large^{8–12}: the energy density in these small-scale fields must far exceed the local kinetic energy density. Here we propose an alternative mechanism of magnetic-field generation in galaxies. We show that a seed field can be generated by the rotation of an aspherical cloud of ionized gas around a central massive black hole. Strong shear flows in the rotating gas amplify this seed field, and a relatively slow galactic wind can transport the field to the outer regions of a galaxy in about 100 million years—a timescale short enough to meet the constraints imposed by the observation of strong fields in very young galaxies^{13,14}.

The first step begins with the initial (catastrophic) collapse of a fragmented protogalactic cloud; we assume that in the context of this collapse, a collapsed core of mass $M \approx 10^6$ solar masses ($M_\odot$) also forms. Thus, we assume that the early epochs of every galaxy resembles an active galactic nucleus, with a black hole

at the centre^{15–19}, surrounded by a radiation-pressure-supported thick accretion torus^{20–22} of almost constant specific angular momentum. Indeed, it is essential for the following argument that the gas in this torus does not rotate along keplerian orbits, and must therefore be supported by pressure forces p_t , within the torus; p_t is then largely derived from the radiation pressure exerted by the ambient photon field.

Assuming that the torus is nearly in hydrostatic equilibrium, we then obtain the force balance relation

$$\frac{\nabla p_t}{\rho} \approx -\nabla\phi + \frac{L^2}{R^3} \hat{e}_R$$

where ρ is the mass density of the ions, ϕ is the gravitational potential due to the central massive object, $L \approx GM/c$ is the specific angular momentum, c is the velocity of light, R is the axial distance from the rotation axis, $\hat{e}_R$ is a unit vector in the radial direction and G is the universal gravitational constant.

To understand the origin of the initial field, we first consider a spherically symmetric ionized-matter distribution in hydrostatic equilibrium (that is, $L=0$). In this case, the electrons drift slightly outwards relative to the photons (because the gravitational scale heights of electrons and protons differ by their inverse mass ratio), but are ultimately restrained by the protons through the Coulomb force. The electrostatic field generated by this electron-proton separation is balanced by the partial electron pressure gradient. Thus, the electric field is irrotational (that is, the gradient of a scalar), and no net current is produced. In contrast, such force balance is not always possible if the matter distribution is not spherically symmetric (that is $L \neq 0$, as expected in the case of the torus just discussed), particularly when the gas is non-barotropic. In this case, the effective gravity (defined by the joint action of gravity and the centrifugal force) cannot be derived from a scalar field^{2,23,24}, and hence the electric field is no longer irrotational; this process is referred to as the

“Biermann battery”^{23,24}. The electric field generated by this process,

$$\mathbf{E}^b = \frac{\nabla p_e}{n_e e} = \frac{\beta m_p}{2c} \frac{\nabla p_i}{\rho}$$

is rotational (p_e , n_e and e are the partial pressure, number density, and the charge of the electrons respectively, m_p is the mass of the proton, and β is the ratio of the gas pressure to the total pressure; the factor of $1/2$ is due to the fact that we assumed pure hydrogen to be the initial constituent). The magnetic field then grows according to the Maxwell equation, $\partial_t \mathbf{B} = c \nabla \times \mathbf{E}^b$, and, assuming that the rotational part of $\mathbf{E}^b$ is of the order $|\mathbf{E}^b|$, we have at this early stage,

$$B_\phi \approx \frac{c |\mathbf{E}^b|}{R_0} t \quad (1)$$

where $R_0 \approx 2GM/c^2$ is the typical length scale of the system and B_ϕ is the toroidal field component. For reasonable equation of state, this toroidal field produced by the ‘battery’ is very small, and needs significant amplification far beyond that obtainable by equation (1) in order to be relevant to observations.

Now, the usual dynamo processes (such as the ‘ α ’-effect) will also operate during this epoch. But recent analytical and numerical calculations^{8–12} have shown that classical dynamo theory, when self-consistently extended to the non linear domain predicts that even very weak magnetic fields (whose strength is far below the local equipartition field B_E , defined by equality between the magnetic energy density and the local energy density) lead to strong suppression of both the generation coefficient (for example, the ‘ α ’-effect) as well as the turbulent diffusivity^{9–12}. Thus, because the magnetic Reynolds number Re_m is extremely large ($Re_m = (2GM/c)/\eta \approx 4.5 \times 10^{18} M_6^{5/8}$ for $\eta \approx 4 \times 10^{12} T^{-3/2} \text{ cm}^2 \text{ s}^{-1}$, where $M_6 \equiv M/10^6 M_\odot$, η is the molecular diffusivity and T is the temperature), standard dynamo processes are extremely ineffective at amplifying these very weak fields. Instead, we can at best expect that during this period, such dynamo processes lead to poloidal fields of order $B_p \approx Re_m^{-1/2} B_E \approx 1.4 \times 10^{-4} M_6^{-13/16} G$, where the equipartition field is given by²² $B_E = (2\pi c^4 m_p / \sigma_T GM)^{1/2} \approx 3 \times 10^5 M_6^{-1/2} G$ (σ_T is the Thomson cross-section). Thus, in this second step, classical dynamo processes lead to poloidal fields comparable in strength to the toroidal fields produced by the Biermann battery (that is, very weak fields, though now in the poloidal direction).

The illustration makes plain that the problem of the origins of cosmic magnetic fields is therefore not with dynamo theory *per se*, but rather with the blind application of its simplified linear version to explain the observed, relatively strong (that is, comparable to B_E) cosmic magnetic fields. If classical magnetic dynamo theory in its standard form will not do, then one must focus on alternative magnetic generation processes. This is exactly what we are now doing.

The final step is to recognize that the strong shear flows close to the central black hole can amplify this poloidal field enormously, leading to linear growth of the toroidal field component, B_ϕ , given by the induction equation

$$\partial_t \mathbf{B} = (\mathbf{B} \cdot \nabla) \mathbf{v} - (\mathbf{v} \cdot \nabla) \mathbf{B} \quad (2)$$

or (where $\mathbf{v}$ is the velocity) to order of magnitude (and taking into account the fact that the toroidal component of the second term on the right-hand side of equation (2) vanishes for axisymmetric ($\partial_\phi = 0$) flows),

$$\tau^{-1} B_\phi \approx B_p \Omega$$

where Ω is the angular velocity of matter close to the black hole, B_p is the poloidal field component and τ is the characteristic time for growth. With $L \sim GM/c$, $\Omega = c^3/GM \approx 0.2 M_6^{-1} \text{ s}^{-1}$, and hence it takes only $\tau \approx 10^{10} M_6^{21/16} \text{ s}$, or less than 10^3 yr for $M_6 = 1$, for B_ϕ to grow to equipartition values of order B_E .

What happens to this magnetic field? If the initial epochs of core galaxy evolution do resemble an active galactic nucleus, then one would expect outward flows (such as jets and bipolar outflows) to remove some of this field to the outer reaches of such galaxies^{1,2}. Similarly, relatively slow galactic winds²⁵ with typical velocities $V_w \approx 10^7 \text{ cm s}^{-1}$, possibly driven by radiation pressure near the galactic core, will also transport these fields to the outer regions of a galaxy over periods of $\sim 10^8 \text{ yr}$, that is, well within the lifetime of such a galaxy. In any case, we expect the strong fields generated during the initial epoch at galactic cores to be dispersed throughout the galactic volume (and beyond). If we simply assume flux conservation during this dispersal process, then we can easily estimate the mean field strength of the dispersed field—independent of the precise spatial characteristics of the wind outflow—by the relation

$$\begin{aligned} \langle B_\phi \rangle_{\text{galactic}} &\approx (R_0/R_{\text{galactic}}) \langle B_\phi \rangle_{\text{core}} \\ &\approx 2 \times 10^{-6} (R_{\text{galactic}}/15 \text{ kpc})^{-1} M_6^{1/2} G \end{aligned}$$

where R_{galactic} is a typical galactic scale length. Comparing to conditions in the solar neighbourhood in our own Galaxy, we note that this estimate roughly agrees with observed values (though substantially stronger fields are seen in some isolated regions of our galaxy). It should be noted that neither observations nor theory give at present any useful guide to the nature of the required wind outflows within these galaxies; this remains the principal uncertainty of our model. Given the present observational evidence for the Milky Way (based on models of tidally-induced star formation rates close to the centre, and comparing the spectral intensity of radiation with models of the accreting flow) which suggest^{15–19} that our Galactic centre does not contain a black hole more massive than $10^6 M_\odot$, our picture seems to be consistent. Note also that in the context of our model, the magnetic field pressure within the galaxy cannot exceed the mean thermal gas pressure, $\langle p_{\text{gas}} \rangle$, as otherwise winds from the galactic core could not have carried such fields outward. This argument imposes an upper limit on the mean magnetic field strength, namely, $\langle B^2 \rangle / 8\pi \lesssim \langle p_{\text{gas}} \rangle$, which becomes relevant to galaxies with core black holes whose mass exceeds $10^6 M_\odot$.

Of much greater potential import for our model are the recent studies^{13,14} of Lyman- α absorption systems in quasar spectra, suggesting the presence of magnetic fields as large as 10^{-5} G during the relatively early stages of galaxy life. These observations are extremely difficult to understand in the context of classical dynamo theory because a rather efficient mechanism is needed²⁶ to explain such large fields so early on in a galaxy’s life. In the context of our model, these observations may be readily understood. Our mechanism is efficient, operates well at large magnetic Reynolds numbers, and produces magnetic fields relatively independent of the actual rotational state of the galaxy in which the black hole source region is embedded (that is, only the shear flows in the immediate vicinity of the core black hole matter as far as the generation process is concerned, not the rotational state of the entire galaxy). We would therefore expect field generation in spirals, as well as elliptical, galaxies as long as a central black hole is present. The efficacy of our mechanism is instead probably limited by the timescale of formation of a black hole in galactic cores; we require this timescale to be very short, $\ll 10^8 \text{ yr}$ (as is theoretically expected²²). Furthermore, we note that because quasar masses are usually higher than we assumed above, that is $M_{\text{quasar}} \approx 10^6 - 10^9 M_\odot$, our model readily accounts for observed large magnetic strengths.

It is well known that a weak poloidal field is able to suppress the growth of toroidal seed magnetic fields produced by the thermal battery effect²⁴ (the so-called ‘back-reaction’ problem). This back-reaction poses a major difficulty in the production of initial stellar magnetic fields by this mechanism. In the present context of proto-galaxies, this is not a difficulty because the field generated by the battery mechanism is continuously removed

from the source region by significant radial and vertical outflows (which are absent inside a star).

How generic is our model? The processes to which we have appealed to construct our magnetic field generation mechanism should operate in all galaxies, provided that there is some angular motion present, and that black-hole formation has taken place. In other words, we have only appealed to processes which are commonly thought to be ubiquitous in very young galaxies. We also note that the same transport process which spreads magnetic fields on galactic scales can also carry fields to intergalactic distances (albeit at a slower rate). As an illustration, at a distance of 1 Mpc from the core of a galaxy producing these fields, our previous scaling suggests magnetic field strengths in

intergalactic regions of the order of $\sim 3 \times 10^{-8} M_6^{1/2} \text{ G}$; this is to be compared to present-day limits on intergalactic fields^{4,27} ($\lesssim 10^{-9} \text{ G}$) as well as intracluster fields²⁸, which is clearly consistent with our estimate if the mass of the central hole is $\sim 10^6 M_\odot$.

Finally, one might be concerned that microwave observations of large-scale cosmic inhomogeneities constrain the magnetic field strength during the periods of early galaxy formation. We recall, however, that unless the magnetic field is quite high ($\sim 10^{-2} \text{ G}$)²⁷ at the very early stages of our Universe, it will not significantly affect the seeding of density perturbations that lead to large-scale structure. However, smaller fields⁴ may suffice to determine the 'shape' of these structures at much later times (E. Kim, A. Olinto and R. R., manuscript in preparation). $\square$

Received 13 September 1993; accepted 7 February 1994.

1. Daly, R. A. & Loeb, A. *Astrophys. J.* **364**, 451–455 (1990).
2. Chakrabarti, S. K. *Mon. Not. R. astr. Soc.* **252**, 246–248 (1991).
3. Ratna, B. *Astrophys. J.* **391**, L1–L4 (1991).
4. Rees, M. J. *Q. J. R. Soc.* **28**, 197–206 (1987).
5. Jaffe, W. J. *Astrophys. J.* **241**, 925–927 (1980).
6. Lawler, J. M. & Dennison, B. *Astrophys. J.* **252**, 81–92 (1982).
7. Krause, M. in *Galactic and Intergalactic Magnetic Fields* (eds Beck, R., Kronberg, P. P. & Wiebeleski, R.) 187–196, (Kluwer, Dordrecht, 1990).
8. Parker, E. N. *Astrophys. J.* **401**, 137–145 (1992).
9. Vainshtein, S. I., Parker, E. N. & Rosner, R. *Astrophys. J.* **404**, 773–780 (1993).
10. Vainshtein, S. I. & Rosner, R. *Astrophys. J.* **376**, 199–203 (1991).
11. Kulsrud, R. M. & Anderson, S. W. *Astrophys. J.* **396**, 606–630 (1992).
12. Tao, L., Cattaneo, F. & Vainshtein, S. I. in *Theory of Solar and Planetary Dynamos* (eds Matthews, P. C. & Rucklidge, A. M.) (NATO ASI, Cambridge Univ. Press, 1993).
13. Kronberg, P. P. & Perry, J. J. *Astrophys. J.* **263**, 518–532 (1982).
14. Wolfe, A. M. in *QSO Absorption Lines: Probing the Universe* (eds Blades, J. C., Turnshek, D. & Norman, C. A.) 297–317 (Cambridge Univ. Press, 1988).

15. Lacy, J. H., Achterman, J. M. & Serabyn, E. *Astrophys. J.* **380**, L71–L74 (1991).
16. McGinn, M. T., Sellgren, K., Becklin, E. E. & Hall, D. N. B. *Astrophys. J.* **338**, 824–840 (1989).
17. Melia, F. *Astrophys. J.* **387**, L25–L28 (1992).
18. Chevalier, R. *Astrophys. J.* **397**, L39–L42 (1992).
19. Ozernoy, L. *Ann. N.Y. Acad. Sci.* **688**, 596–597 (1993).
20. Lynden-Bell, D. *Phys. Scripta* **17**, 185–192 (1978).
21. Paczynski, B. & Wiita, P. J. *Astr. Astrophys.* **88**, 23–32 (1980).
22. Rees, M. J. *A. Rev. Astr. Astrophys.* **22**, 471–506 (1984).
23. Biermann, L. Z. *Nature* **5a**, 65–71 (1950).
24. Mestel, L. & Roxburgh, I. W. *Astrophys. J.* **136**, 615–626 (1962).
25. Ciotti, L., D'Ercole, A., Pellegrini, S. & Renzini, A. *Astrophys. J.* **376**, 380–403 (1991).
26. Heiles, C. A. *Rev. Astr. Astrophys.* **14**, 1–22 (1976).
27. Coles, P. *Comments Astrophys.* **18**, 45–60 (1992).
28. Valentijn, E. A., Perola, G. C. & Jaffe, W. J. *Astr. Astrophys. Suppl. Ser.* **28**, 333–345 (1977).

ACKNOWLEDGEMENTS. This research was supported in part by NASA at the University of Chicago.

Laser action in strongly scattering media

N. M. Lawandy, R. M. Balachandran,
A. S. L. Gomes & E. Sauvain

Division of Engineering and Department of Physics, Brown University, Providence, Rhode Island 02912, USA

THE radiative properties of an atomic or molecular system may be altered significantly in the presence of coherent optical scattering^{1,2}. In the course of investigating the radiative properties of a laser dye dispersed in a strongly scattering medium (a colloidal suspension of titanium dioxide particles), we have found that the emissions from such systems can exhibit spectral and temporal properties characteristic of a multimode laser oscillator, even though the systems contain no external cavity. The threshold excitation energy for laser action is surprisingly low. We suggest that these composite systems might find applications in laser instrumentation and photonics.

The experiments were performed on colloidal solutions containing rhodamine 640 perchlorate dye in methanol and TiO₂ (rutile). The TiO₂ particles were coated with a layer of Al₂O₃ to prevent flocculation and had a mean diameter of 250 nm. A sedimentation time of 14 h was calculated for these colloids, allowing ample time to perform the experiments which took ~ 0.5 h each. The surface area available for adsorption of the dye molecules on the nanoparticles was measured using N₂ adsorption and found to be of the order of $13 \text{ m}^2 \text{ g}^{-1}$. Thus, at a particle density $\rho \approx 10^{10} \text{ cm}^{-3}$, no more than $\sim 1\%$ of the dye molecules of a $2.5 \times 10^{-3} \text{ M}$ dye solution can be accommodated on the nanoparticle surfaces. We can therefore discount the possibility that surface effects play an important role in these systems. The scattering cross-sections at the peak dye emission wavelength of $\sim 617 \text{ nm}$ were calculated using the full Mie solutions and found to be far too small to exhibit resonances capable of supporting laser action in a single particle³. (Such morpholo-

gy-dependent resonances arise due to the reflection of light at the interface of the particle and the surrounding medium⁴.)

The colloidal solutions were optically pumped by linearly polarized 532-nm radiation from a frequency-doubled Nd:YAG laser operating at 1.064 μm . The excitation beam was incident at an angle of 30° with respect to the surface normal of the cuvette containing the colloid. Experiments were performed with either a Q-switched laser which produced single 7-ns-long pulses, or with a Q-switched and mode-locked laser which produced a 125-ns-long train of nine 80-ps-long pulses. The 532-nm radiation has a $50 \mu\text{m}$ small-signal penetration depth in the pure dye solution, making it smaller than the shortest optical scattering lengths used in any of our experiments. The laser spot areas at

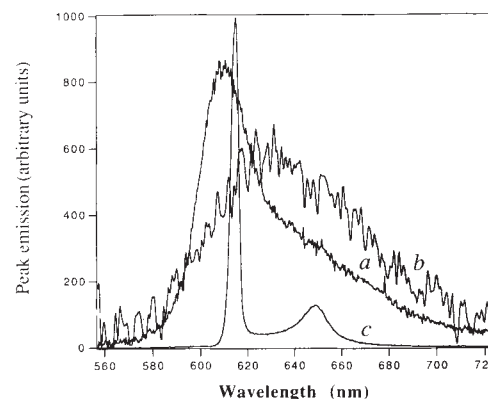


FIG. 1 a, Emission spectrum of a $2.5 \times 10^{-3} \text{ M}$ solution of rhodamine 640 perchlorate in methanol pumped by 3-mJ (7-ns) pulses at 532 nm. b and c, Emission spectrum of the TiO₂ nanoparticle ($2.8 \times 10^{10} \text{ cm}^{-3}$) colloidal dye solution pumped by 2.2- μJ and 3-mJ (7-ns) pulses, respectively. The amplitude of the spectrum in b has been scaled up by a factor of 10, whereas that in c has been scaled down by a factor of 20.

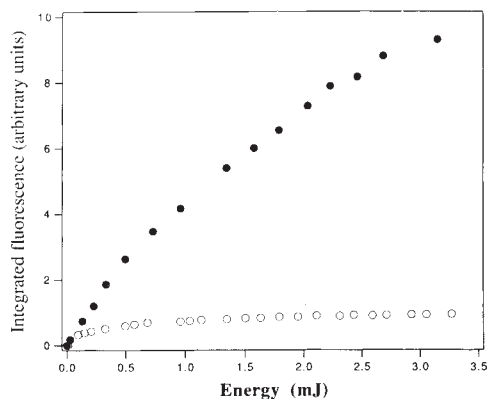


FIG. 2 Wavelength-integrated emission as a function of the pump pulse energy for the pure dye (open circles) and the TiO_2 nanoparticle ($2.8 \times 10^{10} \text{ cm}^{-3}$) colloidal dye solutions (closed circles). The dye concentration is $2.5 \times 10^{-3} \text{ M}$ in both cases.

the cell face were $2.5 \times 10^{-2} \text{ cm}^2$ and $7.85 \times 10^{-3} \text{ cm}^2$ for the 7-ns and 80-ps excitations respectively. All the measurements were performed at a repetition rate of 25 Hz or less, to avoid any dye degradation effects. The maximum energy per pulse for these experiments was $\sim 10 \text{ mJ}$ and 0.12 mJ for the long and short pulses respectively. The emission from the front face of the cell was collected using a lens and sent to an optical multichannel analyser to study the spectrum, and through a monochromator to a fast photodiode and oscilloscope combination to measure the temporal characteristics.

The first series of experiments were performed using the long pulses pumping a $2.5 \times 10^{-3} \text{ M}$ dye solution in 1-cm-thick glass cuvette. The excitation of the pure dye solution resulted in the spectrum in Fig. 1a. This spectrum exhibited a main peak at 610 nm with a shoulder at 640 nm, and remained constant for the entire range of pump pulse energies up to 10 mJ. The wavelength-integrated fluorescence as a function of pump energy showed saturation behaviour with a saturation energy of 0.26 mJ, and is shown by the open circles in Fig. 2. This saturation energy, along with the spot size and pump pulse duration, agrees with a saturation intensity of 0.7 MW cm^{-2} expected from quasi-steady-state rate equation models for optically pumped dyes⁵.

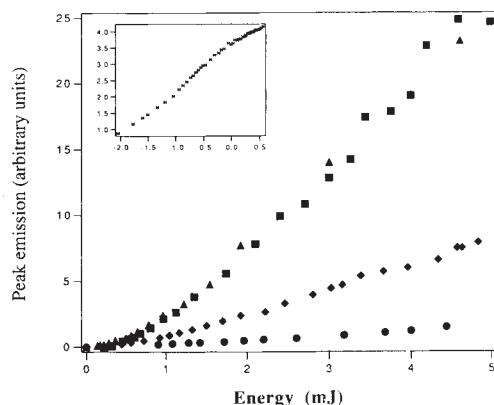


FIG. 3 The emission at the peak wavelength as a function of the pump pulse energy for four different TiO_2 nanoparticle densities. Nanoparticle densities of $1.4 \times 10^9 \text{ cm}^{-3}$, $7.2 \times 10^9 \text{ cm}^{-3}$, $2.8 \times 10^{10} \text{ cm}^{-3}$ and $8.6 \times 10^{11} \text{ cm}^{-3}$ are shown by solid circles, diamonds, squares and triangles respectively. The inset shows the data on a log-log plot (same quantities on the x and y axes) for a nanoparticle density of $2.8 \times 10^{10} \text{ cm}^{-3}$.

Similar optical pumping experiments were performed in the same solution containing $2.8 \times 10^{10} \text{ cm}^{-3}$ of the TiO_2 nanoparticles. The spectrum at the lowest excitation is shown in Fig. 1b and has a linewidth of 70 nm as compared to the 36 nm width of the pure dye solution. This larger emission linewidth is primarily due to scattering enhanced self absorption by the dye which results in re-emission near 650 nm, making the convolved line-shape wider. When the energy of the excitation pulses was increased, the unpolarized emission at 617 nm grew rapidly and narrowed as shown in Fig. 1c. As the pump energy was increased even further, a bichromatic spectrum was observed. The 650-nm emission was only observed in cells thicker than 100 μm and is associated with stimulated emission on a weaker optical transition in the dye molecule. The solid-circle data in Fig. 2 show the wavelength-integrated emission as a function of pump energy for the colloid. Surprisingly, the colloid does not exhibit the strong saturation behaviour observed in the pure dye. Even more striking is the dependence of the 617-nm peak emission on pump energy for various nanoparticle densities shown in Fig. 3. This Figure shows a well defined pump energy threshold for the change in slope in the input-output data for all the particle concentrations studied. When these data are plotted on a logarithmic scale, the result is the characteristic S-shaped curve for laser behaviour shown in the inset of Fig. 3. Analysis of the lineshape data revealed that at the same pump energy where a

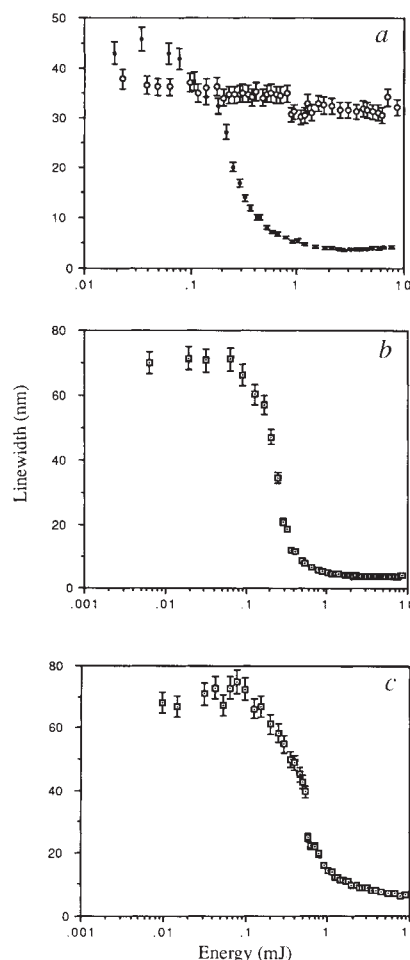


FIG. 4 Emission linewidth as a function of the pump pulse energy for three different TiO_2 nanoparticle densities. a-c correspond to nanoparticle densities of $5.7 \times 10^9 \text{ cm}^{-3}$, $2.8 \times 10^{10} \text{ cm}^{-3}$ and $1.4 \times 10^{11} \text{ cm}^{-3}$, respectively. The open circles in a represent the emission linewidth of the pure dye as a function of the pump pulse energy.

change in slope in the input–output behaviour is observed, the emission linewidth collapses rapidly to 4 nm. Figure 4a–c shows plots of the emission linewidth (full width at half maximum) as a function of the pump pulse energy for three different TiO_2 nanoparticle concentrations. Figure 4a also shows the linewidth of the pure dye as a reference (open circles). These data further corroborate the view that quasi-steady-state laser behaviour is taking place in this system.

Excitation with a train of 80-ps pulses also revealed a threshold behaviour in the temporal characteristics of the colloid. When the pump pulse energy was below that required for the onset of laser action, the peak emission at 617 nm exhibited a long 4-ns decay identical to that observed in the pure dye solutions at all pump energies. In addition, there was a large background signal because the pulses arrived every 13.15 ns, barely allowing the excited molecules enough time to relax. When the pump pulse energy was increased beyond the threshold point, a sharp spike appeared which was shorter than the 300-ps system resolution. Further increasing the energy resulted in only the sharp spike and a complete recovery between pulses in the mode-locked and Q-switched train. These results are summarized in the oscilloscope traces of Fig. 5.

The explanation for the observed laser-like behaviour of the optically pumped colloid is not obvious. One is tempted to think in terms of photon diffusion as providing a kind of non-resonant feedback for the high-gain laser dye. This type of non-resonant

light trapping was first suggested by Letokhov in the late 1960s as a method for producing frequency-stable laser oscillators⁶. One of the main problems with invoking the light diffusion process as the origin of the pseudo-cavity in our experiments is that the effect requires that the smallest dimension of the scattering medium be large compared to the optical scattering length. In the case of our experiments the scattering length at the lasing wavelength was typically of the order of 200 μm , requiring that every dimension of our sample be of the order of several millimetres in order that the diffusion time of photons is a meaningful concept; yet we observed laser-like behaviour in samples just 100 μm thick. In another series of experiments, we found that the linewidth collapse could be observed at cell thicknesses as small as 30 μm or one-sixth the scattering length. These results suggest that the diffusive-type picture in its simplest form is inadequate for explaining our results. We believe that either a two-dimensional version of this process may be a possibility, or that some entirely different mechanism is at play. Experiments with samples that have every dimension smaller than a scattering length, and are index-of-refraction matched at the boundaries, are currently under way. We hope that these future results will help direct us towards a theory for this potentially useful concept in photonics and other applications where high brightness and narrow emission linewidths are essential. $\square$

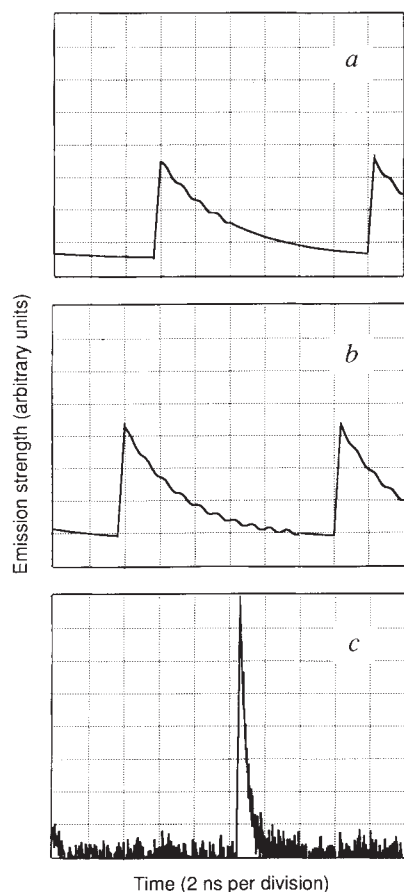


FIG. 5 Temporal emission of the dye and the TiO_2 nanoparticle colloidal dye solution pumped by a train of 80-ps-long pulses. a, Response of the pure dye at the highest pump energy. b and c, Response of the TiO_2 nanoparticle colloidal dye solutions at low (1.2×10^{-2} mJ per pulse) and high (1.2×10^{-1} mJ per pulse) pump energies, respectively. In both cases the dye concentration is 2.5×10^{-3} M and the nanoparticle density is $2.8 \times 10^{19} \text{ cm}^{-3}$.

1. Martorell, J. & Lawandy, N. M. *Phys. Rev. Lett.* **65**, 1877–1880 (1990).
2. Martorell, J. & Lawandy, N. M. *Phys. Rev. Lett.* **66**, 887–890 (1991).
3. Qian, S.-X., Snow, J. B., Tzeng, H.-M. & Chang, R. K. *Science* **231**, 486–488 (1986).
4. Hill, S. C. & Benner, R. E. in *Optical Effects Associated with Small Particles* (eds Barber, P. W. & Chang, R. K.) 3–61 (World Scientific, Singapore, 1988).
5. Schafer, F. P. in *Dye Lasers* (ed. Schafer, F. P.) 1–85 (Springer, Heidelberg, 1977).
6. Ambartsumyan, R. V., Basov, H. G., Kryukov, P. G. & Letokhov, V. S. in *Progress in Quantum Electronics* (eds Sanders, J. H. & Stevens, K. W. H.) 109–185 (Pergamon, Oxford, 1970).

ACKNOWLEDGEMENTS. N.M.L. is grateful to Spectra Science Corporation for supporting this work and to J. M. Calo for providing the N_2 adsorption measurements.

Non-centrosymmetry and second-harmonic generation in Z-type Langmuir–Blodgett films

Geoffrey J. Ashwell, Paul D. Jackson & Wendy A. Crossland

Centre for Molecular Electronics, The Advanced Materials Group, Cranfield University, Cranfield MK43 0AL, UK

THE principal requirement for materials that exhibit second-harmonic generation—frequency-doubling of light—is that they have a non-centrosymmetric structure. For optically active organic materials, this can be achieved by the use of the Langmuir–Blodgett (LB) technique^{1,2}: amphiphilic molecules comprising a hydrophilic head (the chromophore) and a hydrophobic tail are aligned at an air–water interface, and the resulting ordered monolayer is transferred to an appropriate solid substrate. For practical applications, the non-centrosymmetric structure of the individual monolayers needs to be maintained across many such layers, but this has been achieved previously for relatively few optically active molecules^{3–5}. Most amphiphiles pack centrosymmetrically, with the layers adopting a head-to-head and tail-to-tail ‘Y-type’ structure⁶. To enforce a non-centrosymmetric structure, it becomes necessary to intersperse the optically active layers with another material^{7–12}. Here we demonstrate an alternative strategy for achieving macroscopic non-centrosymmetric order in LB multilayers. We find that the addition of a second hydrophobic end-group to the hydrophilic chromophore reduces the tendency of the molecules to invert during deposition, and a head-to-tail ‘Z-type’ structure can be readily obtained for films containing more than 100 layers.

TABLE 1 Physical properties of multilayer LB films of four optically nonlinear dyes

Dye*	Molecular formula	Spectra $\lambda_{\max}(\text{nm})$	Ellipsometry† $I(\text{nm})$	$n_{670\text{nm}}$	$\varphi(\text{deg})$	SHG‡ $\chi_{zzz}^{(2)}(\text{pm V}^{-1})$
I	$(\text{C}_4\text{H}_9)_2\text{N}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{Qn}^+-\text{C}_{18}\text{H}_{37}\text{I}^-$	530	3.0	1.7	36	120
II	$(\text{C}_4\text{H}_9)_2\text{N}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{Py}^+-\text{C}_{18}\text{H}_{37}\text{Br}^-$	460	3.4	1.6	36	50
III	$\text{C}_{17}\text{H}_{35}\text{C}(\text{O})\text{O}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{Py}^+-\text{C}_{22}\text{H}_{45}\text{I}^-$	380	5.2	1.5	32	15
IV	$\text{C}_{22}\text{H}_{45}\text{O}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{Py}^+-\text{C}_{12}\text{H}_{25}\text{Br}^-$	360	5.2	1.5	33	15
V	$\text{C}_{18}\text{H}_{37}\text{O}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{Py}^+-\text{C}_{18}\text{H}_{37}\text{I}^-$	360	5.0	1.5	28	5

* Dye I, E-N-octadecyl-4-(2-(4-dibutylaminophenyl)ethenyl)quinolinium⁺I⁻; dye II, E-N-octadecyl-4-(2-(4-dibutylaminophenyl)ethenyl)pyridinium⁺(C₁₈H₃₇OSO₃)_{1/2}(I⁻)_{1/2}; dye III, E-N-docosyl-4-(2-(4-octadecyloxyphenyl)ethenyl)pyridinium⁺I⁻; dye IV, E-N-dodecyl-4-(2-(4-docosyloxyphenyl)ethenyl)pyridinium⁺Br⁻; dye V, E-N-octadecyl-4-(2-(4-octadecyloxyphenyl)ethenyl)pyridinium⁺I⁻.

† I is the thickness per layer and n is the refractive index.

‡ The tilt angle (φ) of the chromophore relative to the normal was determined from the polarization dependence of the SHG, and in the calculation of $\chi_{zzz}^{(2)}$ it was assumed that $n_{2\omega}=1.5$ (off-resonance) or 1.8 (on-resonance) and $n_{\omega}=1.5$.

Films of $\text{C}_m\text{H}_{2m+1}-\text{Y}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{A}^+-\text{C}_n\text{H}_{2n+1}\text{X}^-$, where Y is O, S, C(O)O or (C_mH_{2m+1})N, A⁺ is a heterocyclic cation and X⁻ is a counterion, were deposited using an LB trough (Nima Technology Ltd, Coventry, UK) with two compartments separated by a fixed surface barrier. The dyes were spread from dilute chloroform solution onto the pure water subphase of one compartment and the floating layer slowly compressed to $\sim 35\text{ mN m}^{-1}$; multilayer films were fabricated by cycling the substrate by means of the air-water interface of the second compartment, under the surface barrier, to deposit on the upstroke at a rate of $50\text{ }\mu\text{m s}^{-1}$. The second-harmonic generation (SHG) from the LB films was measured in transmission using a Nd/YAG laser (λ 1,064 nm) and the data analysed by the general procedure described in ref. 12.

The SHG from films of dyes I to V (see Table 1 for their full names) increases quadratically with the number of layers indicating a Z-type packing arrangement (Fig. 2a). The quinolinium dye (I) has a second-order susceptibility of $\chi_{zzz}^{(2)}=120\text{ pm V}^{-1}$ which compares favourably with the highest value so far obtained for a multilayer film, 180 pm V^{-1} for the related quinolinium zwitterion⁴. Unfortunately, there is significant overlap of the second harmonic ($\lambda_{2\omega}$, 532 nm) and the absorbing charge-transfer band ($\lambda_{\max}$, 530 nm; half-width at half-maximum (HWHM), 78 nm), but the transition may be shifted to shorter wavelengths by using a weaker donor-acceptor combination. Thus, the charge-transfer band of the pyridinium dye with a dibutylamino donor (dye II) is shifted to 460 nm with HWHM = 64 nm, whereas for the $\text{C}_m\text{H}_{2m+1}\text{C}(\text{O})\text{O}$ (dye III) and $\text{C}_m\text{H}_{2m+1}\text{O}$ congeners (dyes IV and V) the band is further shifted to 380 and 360 nm respectively with HWHM $\approx 62\text{ nm}$. For these films there is no evidence of H-aggregation as reported by Schildkraut *et al.*¹³ for a related hemicyanine dye. In addition, the quadratic dependence of the SHG on the number of layers is compelling evidence of long-range structural order.

Langmuir-Blodgett films of $\text{C}_m\text{H}_{2m+1}-\text{O}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{C}_5\text{H}_4\text{N}^+-\text{C}_n\text{H}_{2n+1}\text{I}^-$, where m and n are 12, 18 or 22, are Z-type but the nonlinear optical properties depend on the combination of hydrophobic groups: for example, for dye IV ($m=22$, $n=12$) the second-order susceptibility, $\chi_{zzz}^{(2)}=$

TABLE 2 Film-forming properties of optically nonlinear dyes

Hydrophobic end-groups*

(C ₄ H ₉) ₂ N	Qn ⁺ -C ₁₈ H ₃₇	Z-type
(C ₄ H ₉) ₂ N	Py ⁺ -C ₁₈ H ₃₇	Z-type†
(C ₁₈ H ₃₇)HN	BT ⁺ -C ₁₂ H ₂₅	Z-type
(C ₁₈ H ₃₇)HN	Py ⁺ -C ₁₂ H ₂₅	Z-type
C ₁₇ H ₃₅ C(O)O	Py ⁺ -C ₂₂ H ₄₅	Z-type
C ₂₂ H ₄₅ O	Py ⁺ -C ₁₂ H ₂₅	Z-type

Single hydrophobic tail

(C ₁₈ H ₃₇)HN	Qn ⁺ -CH ₃	Y-type
(C ₁₈ H ₃₇)HN	Py ⁺ -CH ₃	Y-type
C ₁₈ H ₃₇ O	Qn ⁺ -CH ₃	Y-type
C ₁₈ H ₃₇ O	Py ⁺ -CH ₃	Y-type
(CH ₃) ₂ N	Qn ⁺ -C ₁₈ H ₃₇	Y-type
(CH ₃) ₂ N	Py ⁺ -C ₁₈ H ₃₇	Y-type
CH ₃ O	Qn ⁺ -C ₁₈ H ₃₇	Y-type
CH ₃ O	Py ⁺ -C ₁₈ H ₃₇	Y-type
CH ₃ S	Qn ⁺ -C ₁₈ H ₃₇	Y-type
CH ₃ S	Py ⁺ -C ₁₈ H ₃₇	Y-type
CH ₃ C(O)O	Qn ⁺ -C ₁₈ H ₃₇	Y-type
CH ₃ C(O)O	Py ⁺ -C ₁₈ H ₃₇	Y-type

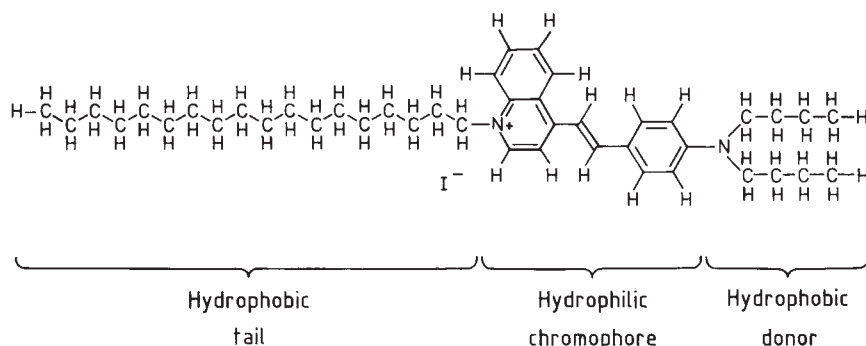
These dyes are of the general formula $\text{C}_m\text{H}_{2m+1}-\text{Y}-\text{C}_6\text{H}_4-\text{CH}=\text{CH}-\text{A}^+-\text{C}_n\text{H}_{2n+1}\text{X}^-$ where A⁺ is a 4-quinolinium (Qn⁺), 4-pyridinium (Py⁺) 2-benzothiazolium (BT⁺) acceptor and X⁻ is a halide.

* For these materials other combinations of hydrophobic chains have been found to give Z-type films.

† (C₁₈H₃₇OSO₃)_{1/2}(I⁻)_{1/2} salt.

15 pm V^{-1} , is similar to that of 5-docosylamino-2-nitropyridine (DCANP)⁵, whereas for dye V ($m=n=18$) the value is reduced to 5 pm V^{-1} . The SHG from films of dye V also increases quadratically (Fig. 2b) but the normalized intensity, $I_{2\omega}/N^2$ where N is the number of Z-type layers, is an order of magnitude lower. This is attributed to the fact that the molecular geometry of dye V is almost symmetrical and thus, within each monolayer there is a tendency to form centrosymmetric defects in which the molecular dipoles are antiparallel.

FIG. 1 Molecular structure of E-N-octadecyl-4-(2-(4-dibutylaminophenyl)ethenyl)-quinolinium⁺I⁻ showing the hydrophilic chromophore and the hydrophobic ends.



Z-type films of optically nonlinear dyes are rare. Based on SHG measurements, only the quinolinium zwitterion has demonstrated an ordered Z-type packing arrangement in thick films (200 layers)⁴. However, quadratic SHG enhancement has been reported by Aktsipetrov *et al.*¹⁴ for an azobenzene dye (five layers) and by Richardson *et al.*¹⁵ from an organoruthenium complex (seven layers). Popovitz-Biro *et al.*¹⁶ have shown similar behaviour for films comprising 2 to 10 layers of an amphiphilic nitroaniline, but the normalized intensity ($I_{2\omega}/N^2$) is 20% of the optimum monolayer signal. Stroeve *et al.*¹⁷ have also reported a weakly increasing second-harmonic intensity to seven layers for a hemicyanine dye, but this may be attributed to a disordered Z-type structure. In addition, Blinov *et al.*^{18,19} have reported pyroelectricity from non-centrosymmetric LB films of azobenzene dyes.

In contrast, our preliminary studies on a series of optically nonlinear dyes with hydrophobic groups at opposite ends of the chromophore have provided thick Z-type structures (Table 2), but if either end of the molecule is substituted by a less hydrophobic group, the interlayer LB packing changes to Y-type; the second-harmonic intensity from the Y-type films is modest when

the number of layers is odd and negligible when even, indicating that the molecular orientation 'flips' during deposition to give a centrosymmetric head-to-head and tail-to-tail array. From this study, it appears that the necessity to rearrange is suppressed when each end of the molecule is hydrophobic, and for nonlinear optical applications it has been shown that non-centrosymmetric Z-type films may be routinely fabricated by careful consideration of the molecular structure. □

Received 27 August 1993; accepted 25 January 1994.

- Allen, S. in *Molecular Electronics* (ed. Ashwell, G. J.) 207–265 (Research Studies Press, Taunton, 1992).
- Bubeck, C. *et al.* *Angew. Chem. int. Edn engl. Adv. Mater.* **3**, 54–58 (1991).
- Ashwell, G. J. *et al.* *J. chem. Soc., Faraday Trans.* **1117–1121** (1990).
- Ashwell, G. J. in *Organic Materials for Nonlinear Optics* (eds Ashwell, G. J. & Bloor, D.) 31–39 Royal Soc. of Chem., Cambridge, 1993).
- Decher, G., Tieke, B., Bossard, C. & Günter, P. *J. chem. Soc., chem. Commun.* 933–934 (1988).
- Peterson, I. R. in *Molecular Electronics* (ed. Ashwell, G. J.) 117–206 (Research Studies Press, Wiley, Taunton, 1992).
- Anderson, B. L. *et al.* *Thin Solid Films* **179**, 413–421 (1989).
- Era, M. *et al.* *J. appl. Phys.* **29**, L2261–L2263 (1990).
- Penner, T. L. *SPIE Int. Soc. opt. Engng* **1360**, 377–386 (1991).
- Hsiung, H., Rodriguez-Parada, J. & Bekerbauer, R. *Chem. phys. Lett.* **182**, 88–92 (1991).
- Ashwell, G. J., Dawnay, E. J. C., Kuczyński, A. P. & Martin, P. J. *SPIE Int. Soc. opt. Engng* **1361**, 589–598 (1991).
- Ashwell, G. J. *et al.* *Proc. R. Soc. A* **445**, 1–14 (1994).
- Schildkraut, J. S., Penner, T. L., Willand, C. S. & Ulman, A. *Opt. Lett.* **13**, 134–136 (1988).
- Aktsipetrov, O. A. *et al.* *Soviet Tech. Phys. Lett.* **11**, 249–251 (1985).
- Richardson, T., Roberts, G. G., Polywka, M. E. C. & Davies, S. G. *Thin Solid Films* **160**, 231–239 (1988).
- Popovitz-Biro, R. *et al.* *J. Am. chem. Soc.* **110**, 2672–2674 (1988).
- Stroeve, P., Saperstein, D. D. & Rabolt, J. F. *J. chem. Phys.* **92**, 6958–6967 (1990).
- Blinov, L. M., Davydova, N. N., Lazarev, V. V. & Yudin, S. G., *Soviet Phys. Solid State* **24**, 1523–1525 (1982).
- Blinov, L. M., Mikhnev, L. V., Sokolova, E. G. & Yudin, S. G., *Soviet Tech. Phys. Lett.* **9**, 640–641 (1983).

ACKNOWLEDGEMENTS. G.J.A. thanks the UK SERC for support of the nonlinear optics programme, and P.D.J. thanks the UK SERC and British Gas p.l.c. for a CASE studentship.

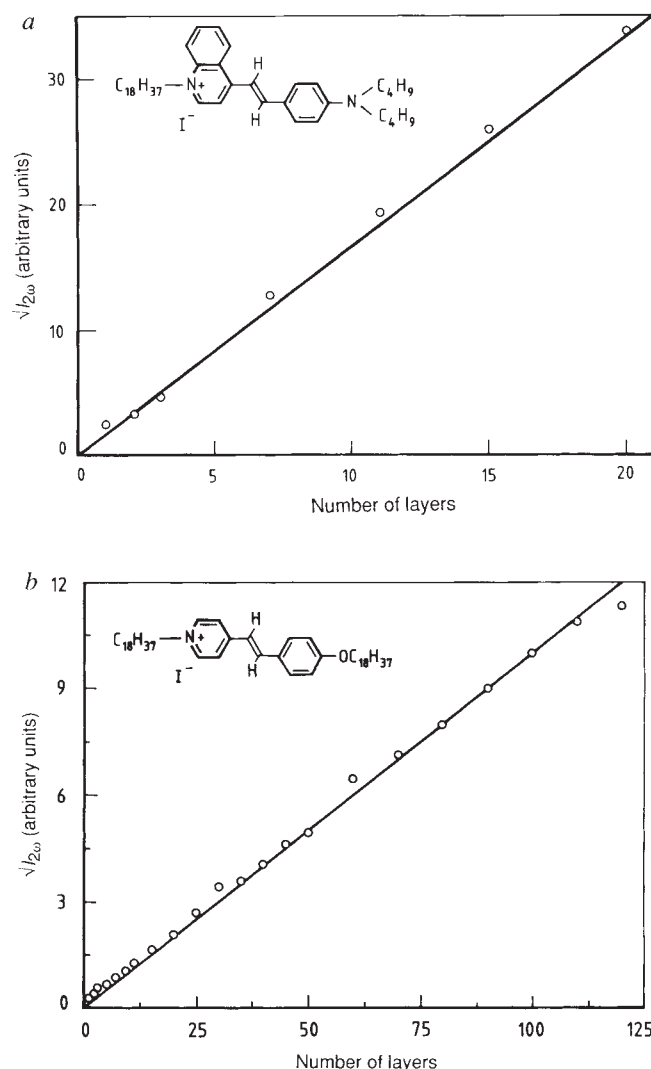


FIG. 2 a, b, Quadratic dependence of the second-harmonic intensity with the number of LB layers of E-N-octadecyl-4-(2-(4-dibutylamino-phenyl)ethenyl)quinolinium⁺I⁻ (a) and E-N-octadecyl-4-(2-(4-octadecyloxyphenyl)ethenyl)pyridinium⁺I⁻ (b). The film thickness from the ellipsometry data is consistent with the molecule being linear with the hydrophobic chains pointing in opposite directions.

Spontaneous chiral symmetry breaking by achiral molecules in a Langmuir–Blodgett film

R. Viswanathan, J. A. Zasadzinski* & D. K. Schwartz

Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA
Department of Chemistry, University of California, Los Angeles, California 90024, USA

THE spontaneous breaking of symmetry is responsible for many physical phenomena, including the mass differences of elementary particles and phase transitions in condensed-matter systems. The breaking of mirror symmetry leads to chirality. In general, chiral effects in condensed-matter systems, such as optical activity, are associated with chiral molecular structures. But several recent observations^{1–3} of domain formation in thin organic films of achiral molecules have suggested that spontaneous separation into regions of differing chirality may occur in these systems. Eckhardt *et al.*⁴, meanwhile, have reported the spontaneous resolution of chiral molecules in Langmuir–Blodgett (LB) films. Here we present images of LB films of calcium arachidate obtained with the atomic force microscope⁵, which suggest that these achiral molecules can separate spontaneously into lattices with chiral packing of opposite handedness. We suggest that this symmetry breaking is driven by the interplay between the packing constraints imposed by the alkane tails and the molecular-area requirements set by the calcium ions^{6–13}.

* To whom correspondence should be addressed.

In addition to being model systems for the study of molecular interactions that lead to self-assembly, LB films have applications in electronics, nonlinear optics, cell membrane models and biosensors^{14,15}. Although the structure and symmetry of LB films have been studied by a variety of techniques, the atomic force microscope (AFM) has increasingly been used to image molecular structure and defects^{6,12,16,19}. The AFM has the advantage of probing only surface topography, and the capability to resolve individual molecules as well as defects such as dislocations and domain boundaries in the surface layer^{6-12,17}.

Figure 1a shows a typical unenhanced 10×10 nm AFM image of the methyl end of trilayers of calcium arachidate (CaA) on mica with the two-dimensional Fourier transform (FT) inset. Although the molecular rows (and some periodicities along the rows) are easily visible in the image, the details of the molecular lattice are much more readily determined from the FT. The six brightest spots in the FT form a skewed hexagonal pattern and correspond to three distinctly different lattice spacings in the film of (A) 0.442 nm, (B) 0.576 nm, and (C) 0.686 nm. The FT has no axis of mirror symmetry; hence the real-space CaA lattice also has no mirror symmetry and is chiral, even though the constituent molecules are achiral. Figure 1b and c are representative FTs of images (not shown) of two different crystalline domains in the same CaA LB film. Figure 1c is similar to the inset in Fig. 1a and has the same progression (clockwise) ABC of lattice spacings. Figure 1b is the mirror image of Fig. 1c and has the same lattice spacings but the progression (clockwise) is ACB; the two FTs cannot be superimposed by any two-dimensional rotation or translation. This shows that the CaA film consists of domains of two distinct lattice structures that are mirror images of each other, that is, that the achiral CaA molecules have undergone a spontaneous chiral symmetry breaking. The crystallinity of the mica substrate had no effect on this

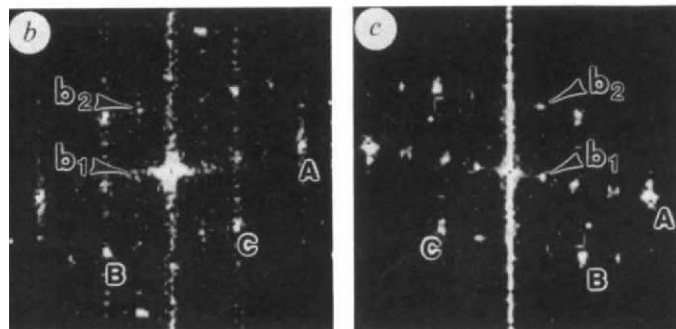
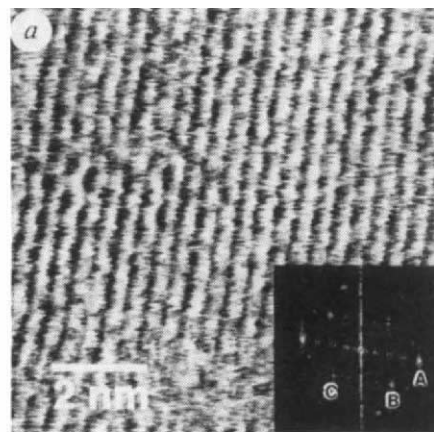
ordering as monolayers of CaA showed a smooth and featureless surface without any lattice structure, similar to cadmium arachidate films^{8,11}.

There are several possibilities for chiral symmetry breaking in thin films of achiral molecules: (1) a hexatic phase with the tilt direction between nearest and next-nearest directions (smectic L phase²⁰), (2) inequivalent mirror-image molecular packings, and (3) chiral resolution of a racemic mixture of two opposite enantiomers²¹. Case (3) is a particular instance of case (2)⁴. Our AFM images of CaA LB films are the first molecular-scale resolution of case (2), a more general example than that reported by Eckhardt *et al.*⁴, who also observed mirror-image lattice structures but ascribed the chirality to a resolution of a racemic mixture of chiral molecules. In fact, our results show that resolution of a racemic mixture can not be safely inferred from the existence of a chiral lattice structure. Although case (1) has been inferred from long-length-scale striped patterns in Langmuir monolayers^{1,2} and thermotropic smectic liquid crystals³, a smectic L phase has never been identified at the molecular level in a system where the constituent molecules were achiral²⁰.

From the information available from the AFM images and FTs we can propose a lattice structure to explain the origin of the chirality. The additional reflections located within the skewed hexagon pattern show that the unit cell of the lattice consists of more than one molecule. A statistical analysis¹¹ was used to determine the best fit to the reciprocal lattice vectors (r.l.v.), $\mathbf{b}_1$, ($\mathbf{b}_1$ and $\mathbf{b}_2$ are labelled in Fig. 1b, c). The positions of the r.l.v. are enough to determine the unit cell dimensions, 1.86 ± 0.01 nm $\times$ 0.894 ± 0.006 nm with an angle of $71.3 \pm 0.5^\circ$ between them (Fig. 1b, c) for a molecular area of 19.7 ± 0.3 Å², assuming an 8-molecule unit cell. A height profile along the [10] direction (Fig. 2a) shows a regular sawtooth pattern with a repeat every four molecular rows; however, we could not accurately deter-

FIG. 1 a, Unprocessed $10 \text{ nm} \times 10 \text{ nm}$ atomic force microscope (AFM) image of trilayer LB films of calcium arachidate and (inset) its two-dimensional Fourier transform (FT). The molecular rows in the film are clearly visible, along with evidence of periodicity along the rows. The six strongest reflections in the FT (inset) form a skewed hexagonal pattern with three distinct lattice repeats of (A) 0.442 nm, (B) 0.576 nm, and (C) 0.686 nm. The sharpness of the reflections indicates long-range positional order in the films. The vertical and horizontal streaks in the FTs are the result of noise from the raster pattern of the AFM. b, c, Mirror-image Fourier transforms of two different domains on a single LB film of calcium arachidate (CaA). The sequence of lattice spacings in c is the same as in a, ABC, while the sequence in b is ACB, although the lattice dimensions are identical. The two FTs are not superimposable by any rotation within the plane of the image. $\mathbf{b}_1$ and $\mathbf{b}_2$ show the reciprocal lattice vectors corresponding to the 4×2 unit cell of CaA (see Fig. 4).

METHODS. Arachidic acid ($\text{CH}_3(\text{CH}_2)_{18}\text{COOH}$) was spread from chloroform (1.8 mg ml^{-1}) onto an aqueous subphase containing 0.5 mM CaCl_2 adjusted to pH 9 with NaOH (Arachidic acid, 99%, Aldrich; chloroform, Fisher spectroanalysed; water, Milli-Q, Millipore, Bedford, Massachusetts; CaCl_2 , 99.9%, Aldrich; NaOH, 99%, Aldrich.) Calcium arachidate films reach full complexation³⁰ at pH >8 , hence the adjustment to pH 9. However, we noticed no significant difference for films made at pH = 6.5 and those made at pH >8 . LB multilayers were deposited on freshly cleaved mica substrates that were previously cleaned in ethanol. Isotherms and film deposition were done on a NIMA (NIMA Technology, Coventry, UK) trough at $22.0 \pm 0.5^\circ\text{C}$ at a surface pressure of $30 \pm 0.1 \text{ mN m}^{-1}$. Film transfer was by vertical dipping at approximately 2.0 mm min^{-1} ; transfer ratios were approximately unity. The AFM measurements were performed with a Nanoscope III FM (Digital Instruments, Santa Barbara, California) at ambient temperature. A $1 \mu\text{m} \times 1 \mu\text{m}$ scan head and a silicon nitride tip on a cantilever with a spring constant of 0.12 N m^{-1} were used. The best molecular resolution was achieved in the 'force' mode, that is, scanning the tip at constant height and measuring spring deflection. Lighter areas in the images are higher. Typical forces were 10^{-8} N .



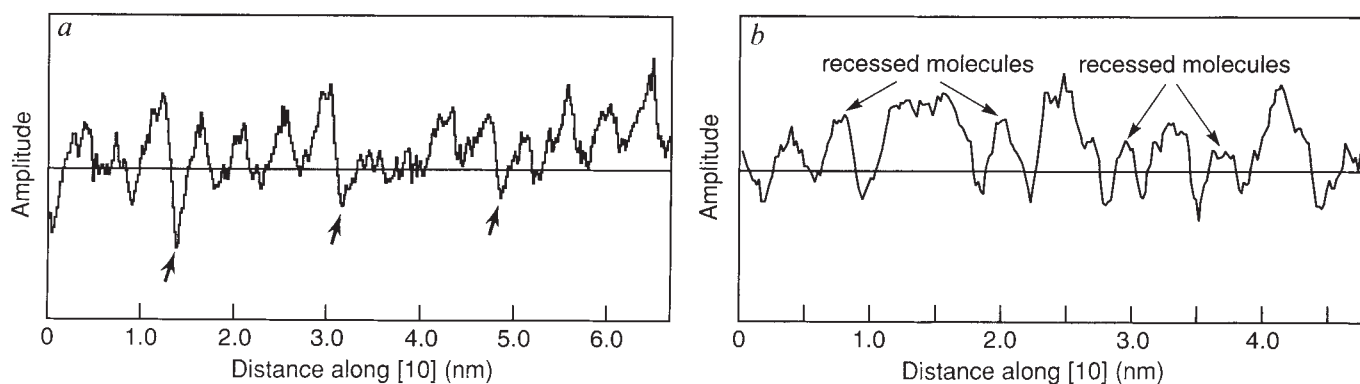


FIG. 2 *a*, Height profile along the [10] direction in the CaA LB film (Fig. 1) showing the regular sawtooth pattern with a repeat every four molecular rows, highlighted by the arrows. The amplitude, which was $\sim 1\text{--}2\text{ \AA}$, could not be accurately determined due to difficulty in calibrat-

ing the z piezoelectric scanner. *b*, Height profile along the [01] direction showing the molecular rows alternating up and down (similar to the 2×2 structure of barium arachidate⁶). The data was smoothed by box-car averaging.

mine the absolute amplitude. A height profile along the [01] direction (Fig. 2*b*) shows that the molecular rows alternate up and down (similar to the 2×2 structure of barium arachidate⁶).

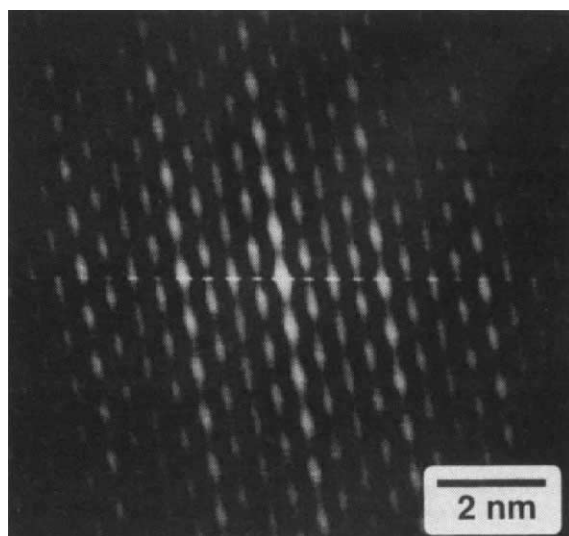
From the autocorrelation function of the images (see Fig. 3), we can extract the local packing dimensions within the unit cell ($0.46\text{ nm} \times 0.45\text{ nm}$ with an angle of 79° —see Fig. 4*a*), confirm the up and down alternation along the rows, and note the increased intensity every fourth row. This leads us to construct a 4×2 unit cell for the CaA lattice. By analogy with barium arachidate (BaA) films⁶, the unit cell is constructed by inverting every fourth local cell (in the 3×1 structure of BaA⁶ every third local cell is inverted). Combined with the regular height modulation (Fig. 2*a*), this suggests that there is a stacking fault every four rows in the [10] direction. Figure 4*b* shows a structural model based on a locally triclinic hydrocarbon packing (Kitaigorodskii's $T[1/2\ 0]$), with a stacking fault after four rows that is quantitatively consistent with our data. Asymmetric stacking in molecular films with stacking faults in every fourth row were also reported for monolayers of liquid-crystal-forming alkyl cyanobiphenyl molecules on graphite²².

The triclinic packing is inherently chiral, leading to both left-handed and right-handed packings that cannot be superimposed (Fig. 4*b*). (The molecules tilt along a direction between nearest

and next-nearest neighbours, which also adds to the chirality.) The stacking fault results in a regular sawtooth pattern along the [10] direction with an estimated amplitude of $\sim 0.15\text{--}0.2\text{ nm}$. The overall effect of the stacking fault is to increase the average tilt of the molecules, from the standard $T[1/2, 0]$ packing subcell tilt of 19° to a 26° tilt. The 26° tilt is consistent with the measured bilayer thickness of 5.0 nm (the extended length of an arachidic acid molecule is $\sim 5.5\text{--}5.6\text{ nm}$: for a tilt of 26° , the bilayer thickness is $5.6 \times \cos 26^\circ = 5.0\text{ nm}$). The observed tilt is also consistent with Fourier transform infrared and near-edge X-ray absorption fine structure (NEXAFS) measurements that show CaA monolayers to be tilted^{23,24}.

The stacking fault is necessary because the all *trans* chains of the alkane portion of the fatty acid can only close-pack at certain discrete tilt angles, resulting in discrete values for the area per molecule¹³. These discrete areas per alkane chain are not necessarily consistent with the density preferred by the counterion. Preliminary calculations (D.K.S., C. M. Knobler, R.V., J.A.Z., manuscript in preparation) show that it is energetically favourable for the lattice to incorporate a series of regular stacking faults while maintaining the close-packed triclinic structure with a tilt of 19° , rather than to adopt a less close-packed structure with a homogeneous 26° tilt. As a consequence of these conflicting density requirements, the 4×2 unit cell of CaA has a locally

FIG. 3 Autocorrelation function of Fig. 1*a*. Note the alternating vertical displacement of the terminal methylene group in the autocorrelation function, shown by the dark holes in between the bright features along each row of molecules. The stacking fault every four rows is also noticeable as every fourth row is highlighted. From the autocorrelation function, the local cell dimensions, as well as the vertical modulations are readily measured. The chequered nature of the packings is verified.



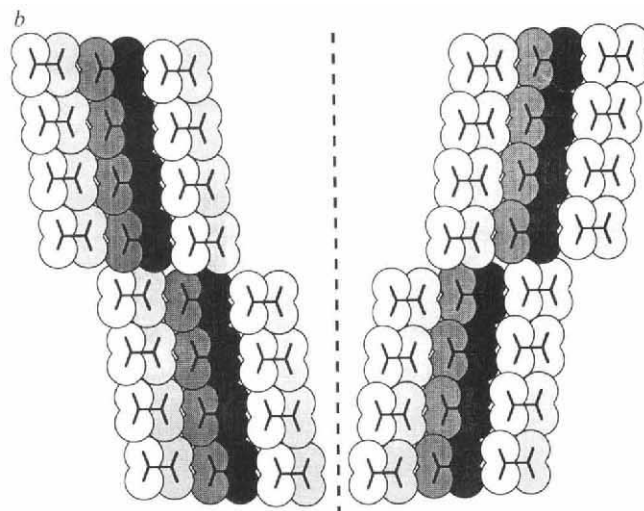
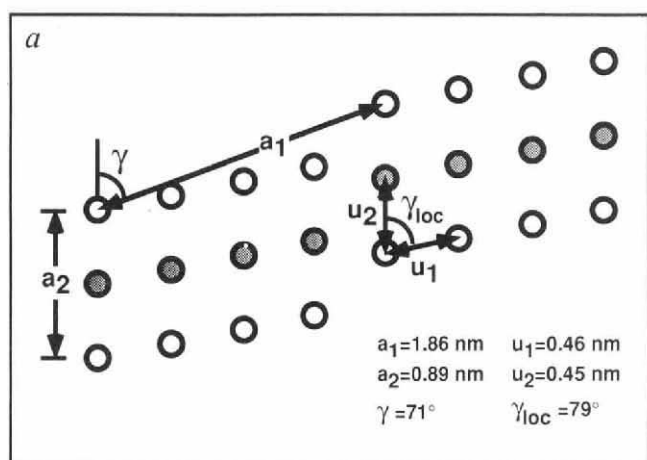


FIG. 4 *a*, Unit cell diagram of a 4×2 structure showing the eight-molecule unit cell dimensions a_1 and a_2 (determined from the position of the Fourier reflections b_1 and b_2) as well as the local molecular packing dimensions within the unit cell u_1 and u_2 determined from the autocorrelation function (Fig. 3). The unit cell can be constructed by inverting every fourth local cell, corresponding to a specific type of stacking fault, as in barium arachidate films⁶. The circles represent the position of the terminal methyl groups, with the shaded circles being displaced vertically by a chain repeat distance (2.54 Å) relative to the lighter circles. *b*, Structural model, based on local triclinic hydrocarbon packing (Kitaigorodskii's $T[1/2 \ 0]$)¹³ with a stacking fault after four rows,

that is quantitatively consistent with our data. The view is along the chain axis of the molecules. The triclinic structure is inherently chiral and we show the mirror plane (dotted line) and the right- and left-handed structures. The black lines represent the projected hydrocarbon backbone of the chain, the circles represent the radii of the hydrogen atoms. The shaded and unshaded circles represent the hydrogen atoms belonging to the even and odd carbons, respectively. The 4×2 unit cell can be constructed by inverting every fourth local cell, corresponding to a regular pattern of stacking faults that causes a jump in the structure both vertically and in the plane (see *a*). As in the AFM images, lighter (unshaded) molecules are higher, darker (shaded) molecules lower.

triclinic structure (to maximize the van der Waals contact between the alkane chains of the fatty acid), but incorporates a regular stacking fault and the up/down displacement present in the 3×1 and 2×2 structures, respectively, of BaA films⁶. These defects are necessary to match the alkane packing density with the area per molecule set by the Ca^{2+} ion. The similarity in lattice structure between Ba^{2+} and Ca^{2+} is not surprising as both are alkaline-earth metals and induce similar effects in Langmuir monolayers^{25,26}. From this and our previous work^{6,12}, it is clear that the counterion sets the lateral density or area per molecule of the fatty acid lattice, and that the alkane chains then pack in such a way as to maximize contact. The simplest correlation, which we observe here, is the inverse correspondence between the molecular area of the multilayer films and the Pauling electronegativity of the cations^{7,25–28}. The local bond angles between the metal and oxygen or type of bond (ionic/covalent) also appear to be important, however, as Ba^6 , Ca and Mg (not

shown) fatty-acid salts all pack into chiral triclinic 'herringbone' lattices, whereas Cd , Mn and Pb salts pack into rectangular lattices, even though the molecular areas of Ca and Mn salts are very similar⁷.

Spontaneous separation has been observed⁴ with AFM in monolayers of racemic mixtures of chiral tetracyclic alcohols in which the symmetry breaking was ascribed to the chirality at the molecular level, similar to three-dimensional resolution of enantiomers by crystallization as shown by Pasteur²⁹. In our present experiments with CaA, the chirality is induced by the novel lattice structure forced on the achiral fatty acids by the competing requirements of the calcium ion molecular area (and likely bond angles) and the allowable close-packing arrangements of alkane chains. Future experiments will show if this spontaneous chiral symmetry breaking also leads to long-length-scale pattern formation similar to those found in Langmuir monolayers^{1,2} and smectic liquid crystals³. □

Received 30 September 1993; accepted 27 January 1994.

- Qui, X., Ruiz-Garcia, J., Stine, K. J., Knobler, C. M. & Selinger, J. V. *Phys. Rev. Lett.* **67**, 703–706 (1991).
- Qui, X., Ruiz-Garcia, J. & Knobler, C. M. *Mat. Res. Soc. Symp. Proc.* **237**, 263–270 (1992).
- MacLennan, J. & Seul, M. *Phys. Rev. Lett.* **69**, 2082–2085 (1992).
- Eckhardt, C. J. *et al. Nature* **362**, 614–616 (1993).
- Binnig, G., Quate, C. F. & Gerber, Ch. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Schwartz, D. K., Viswanathan, R. & Zasadzinski, J. A. *Phys. Rev. Lett.* **70**, 1267–1270 (1993).
- Schwartz, D. K., Viswanathan, R. & Zasadzinski, J. A. *N. J. Am. chem. Soc.* **115**, 7374–7380 (1993).
- Schwartz, D. K., Garnaes, J., Viswanathan, R. & Zasadzinski, J. A. *Science* **257**, 508–511 (1992).
- Garnaes, J., Schwartz, D. K., Viswanathan, R. & Zasadzinski, J. A. *Nature* **357**, 54–57 (1992).
- Schwartz, D. K., Viswanathan, R. & Zasadzinski, J. A. *J. phys. Chem.* **96**, 10444–10447 (1992).
- Schwartz, D. K., Garnaes, J., Viswanathan, R., Chiruvolu, S. & Zasadzinski, J. A. *Phys. Rev. E* **47**, 452–460 (1993).
- Viswanathan, R., Zasadzinski, J. A. & Schwartz, D. K. *Science* **261**, 449–452 (1993).
- Kitaigorodskii, A. I. *Organic Chemical Crystallography* (Consultant Bureau, New York, 1961).
- Roberts, G. G. *Langmuir-Blodgett Films* (Plenum, New York, 1990).
- Ulman, A. *An Introduction to Ultrathin Organic Films* (Academic, San Diego, 1991).

- Marti, O. *Science* **239**, 50–53 (1988).
- Bourdieu, L., Silberzan, P. & Chatenay, D. *Phys. Rev. Lett.* **67**, 2029–2032 (1991).
- Meyer, E. *et al. Nature* **349**, 398–400 (1991).
- Zasadzinski, J. A. N. *et al. Biophys. J.* **59**, 755–760 (1991).
- Sirota, E. B., Smith, G. S., Safinya, C. R., Plano, R. J. & Clark, N. A. *Science* **242**, 1406–1409 (1988).
- Selinger, J. V., Wang, Z.-G., Bruinsma, R. F. & Knobler, C. M. *Phys. Rev. Lett.* **70**, 1139–1142 (1993).
- Smith, D. P. E., Hörber, J. K. H., Binnig, G. & Nejh, H. *Nature* **344**, 641–644 (1990).
- Outka, D. A., Stöhr, J., Rabe, J. P., Swalen, J. D. & Rotermund, H. H. *Phys. Rev. Lett.* **59**, 1321–1324 (1987).
- Ahn, D. J. & Franses, E. I. *J. phys. Chem.* **96**, 9952–9959 (1992).
- Yazdani, M., Yu, H. & Zografi, G. *Langmuir* **6**, 1093–1098 (1990).
- Yazdani, M., Yu, H. & Zografi, G. *Langmuir* **8**, 630–636 (1992).
- Shnidman, Y., Ullman, A. & Eilers, J. E. *Langmuir* **9**, 1071–1081 (1993).
- Pauling, L. *The Nature of the Chemical Bond* 43 (Cornell Univ. Press, Ithaca, New York, 1960).
- Pasteur, L. C. r. *hebd. Seanc. Acad. Sci., Paris* **26**, 535–539 (1848).
- Kobayashi, K., Takoaka, K. & Ochiai, S. *Thin Solid Films* **159**, 267–273 (1988).

ACKNOWLEDGEMENTS. We dedicate this letter to the memory of A. Weisenhorn. We thank C. Knobler, M. Marron and P. Hansma for discussion, and F. Grunfeld for help with the trough and software.

Voltage-dependent ordering of water molecules at an electrode–electrolyte interface

Michael F. Toney*, Jason N. Howard*[†],
 Jocelyn Richer*[‡], Gary L. Borges*,
 Joseph G. Gordon*, Owen R. Melroy*,
 David G. Wiesler*[‡], Dennis Yee[§]
 & Larry B. Sorensen[§]

* IBM Research Division, Almaden Research Center, San Jose, California 95120, USA

[†] Present addresses: Motorola, Inc., 8000 West Sunrise Boulevard, Fort Lauderdale, Florida 33322, USA (J.N.H.); AECL-Research, Chalk River, Ontario K0J 1J0, Canada (J.R.)

[‡] National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

[§] Department of Physics FM-15, University of Washington, Seattle, Washington 98195, USA

THE arrangement of water molecules at charged, aqueous interfaces is an important question in electrochemistry, geochemistry and biology. Theoretical studies^{1–11} suggest that the molecules become arranged in several layers adjacent to a solid interface, with densities similar to that in the bulk, and that the molecules in the first layer are reoriented from oxygen-up to oxygen-down as the electrode charge changes from negative to positive. Few of these predictions have been verified experimentally^{12–16}, however. Using X-ray scattering, we have measured the water density profile perpendicular to a silver (111) surface at two applied voltages. We find that the water molecules are ordered in layers extending about three molecular diameters from the electrode, and that the spacing between the electrode and first water layer indicates an oxygen-up (oxygen-down) average orientation for negative (positive) charge. Contrary to current models, however, we find that the first layer has a far greater density than that in bulk water. This implies that the hydrogen-bonding network is disrupted in this layer, and that the properties of the water in the layer are likely to be very different from those in the bulk.

Statistical models and computer simulations^{1–11} predict that water is ordered in layers extending several molecular diameters from an interface, and that this ordering results from steric forces. There have been suggestions^{3,4,8,9} that, driven by hydrogen bonding, the first layer adopts a solid-like structure similar to ice-I, where the oxygen atoms form puckered hexagons linked by hydrogen bonds. As the electrode charge (or applied voltage) changes from negative to positive, electrostatic considerations

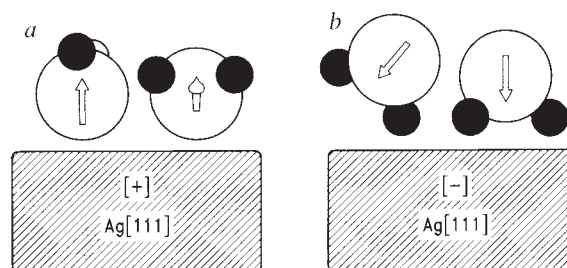


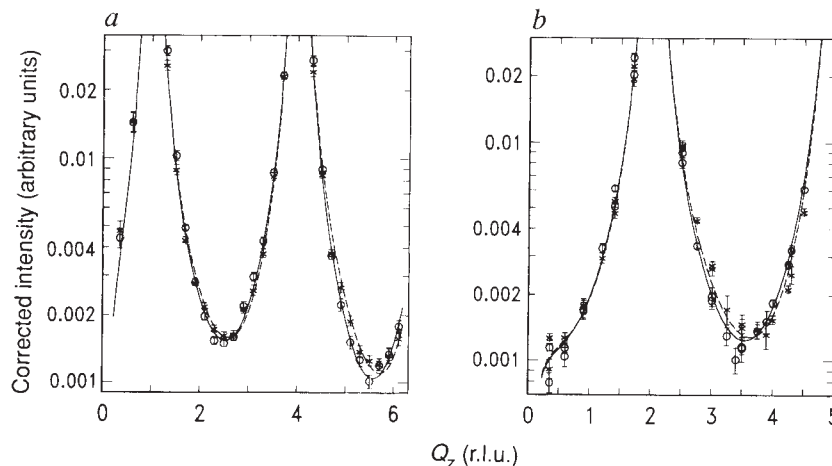
FIG. 1 Possible orientations for water near a surface. *a*, Positively charged surface. *b*, Negatively charged surface. These orientations provide the most favourable electrostatic interactions. The arrows show the directions of the dipole moments. Note that the oxygen atoms are farther from the electrode surface in *b*.

suggest that in the first layer the average molecular orientation changes from oxygen-up to oxygen-down^{1,2,10,11}. This would result in a decrease in the spacing between the oxygen atoms and the electrode (Fig. 1), which is augmented by the movement of the metal electron-density profile towards the electrode with increasing voltage⁶. Such a reorientation need not significantly affect the predicted ice-like structure⁴.

Despite these extensive theoretical studies, direct experimental measurements are sparse. Evidence for the layering comes from mica/electrolyte/mica surface-force experiments¹⁶, which show damped, oscillatory force–separation curves, and from tunnel-junction conductance measurements between mercury droplets¹². The existence of a first water layer with a bulk-like density has been interpreted from X-ray scattering experiments on gold (111) (ref. 15), but these are complicated by reconstruction of the gold surface. Some spectroscopic studies have been interpreted as consistent with the expected first-layer reorientation¹⁴, but others seem inconsistent with this¹³.

To test current ideas, we have used *in situ* X-ray scattering^{17,18} to investigate the water distribution perpendicular to an aqueous–metal interface. An interface at a single crystal creates X-ray scattering that is additional to the Bragg peaks of the crystal. This truncation rod consists of a ridge of intensity that runs perpendicular to the interface and through Bragg points^{17,19}. For the specular truncation rod, the scattering vector **Q** (difference between incident and scattered X-rays) is normal to the interface and the intensity depends on the structure of both the electrolyte and electrode surface perpendicular to the interface. For non-specular truncation rods (where **Q** has a component along the electrode surface) the intensity only depends on the electrode surface structure, because parallel to the interface the water is

FIG. 2 Corrected intensities for the (10) rod (*a*) and the (01) rod (*b*) of Ag(111) at +0.52 V (stars) and –0.23 V (open circles) in 0.1 M NaF. These are the integrated intensities corrected for sample area, resolution function and Ag atomic form factor. They are plotted as a function of Q_z , the component of **Q** normal to the Ag(111) surface (for example, *z* is normal to the surface). The dashed and solid line show the best fits to the data at 0.52 V and –0.23 V, respectively, where the surface roughness and the spacing between the top and second Ag(111) planes were allowed to vary. The abscissa is in reciprocal lattice units (r.l.u.) so that the Bragg peaks occur at integers (1 r.l.u. = 0.88 Å^{–1}).



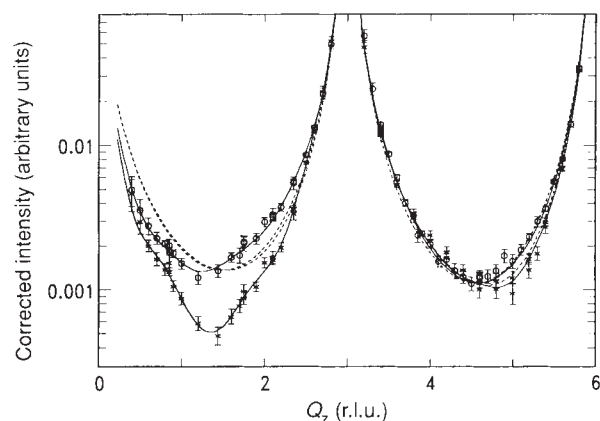


FIG. 3 Integrated intensities of the specular rod for Ag(111) at +0.52 V (stars) and -0.23 V (circles) in 0.1 M NaF, corrected as in Fig. 2. The dashed lines are for bare Ag(111), using the surface roughness and top-second plane spacing obtained from the fits in Fig. 2, but with no contribution from the water distribution. These lines are slightly different at the two voltages, because of the different top-second plane spacings discussed in the text. The solid lines are the best fits with the contribution from the water.

weakly ordered. Thus, measurement of both the specular and non-specular rods permit an independent determination of the surface structure and the water distribution.

The Ag(111) single crystals were prepared using a chromic acid etch²⁰, and the NaF electrolytes were prepared from ultra-pure reagents (Johnson Matthey) and HPLC deionized water. NaF was chosen because NaF/Ag(111) serves as a model system where the ions do not appreciably adsorb on the electrode^{21,22}. Cleanliness of the crystal and electrolytes was checked by comparing the voltammetry and differential capacitance with previous reports^{21,22}. Data are reported for +0.52 and -0.23 V relative to the potential of zero charge (p.z.c.) on the electrode (-0.67 V *vs* Ag/AgCl); these voltages correspond to surface charge densities of ~ 25 and $-10 \mu\text{C cm}^{-2}$, respectively²².

The non-specular truncation-rod intensities at the two voltages are shown by the stars and circles in Fig. 2. The similarity of the profiles shows that the Ag(111) surface does not substantially change with voltage (no reconstruction). But least-square fitting of these data shows that at +0.52 V the spacing between the top and second Ag(111) planes is reduced by 0.03 Å relative to the bulk spacing, whereas at -0.23 V, the spacing is the same as in bulk. The electrostrictive effect has not been included in present-day models of electrified interfaces; a possible explanation will be given elsewhere (M.F.T. *et al.*, manuscript in preparation).

The water distribution perpendicular to the Ag(111) electrode was determined from the specular truncation rod data, shown in Fig. 3 for 0.1 M NaF. The dashed lines are the expected intensities for a bare Ag(111) surface (no contribution from the water). These profiles fail to describe the data, and unlike the non-specular data, the measured intensities change substantially with voltage, which shows that the water distribution changes considerably with voltage. This distribution is quantified with fits to the data where the water density adjacent to the electrode is allowed to vary, and the best fits are shown by the solid lines. Figure 4 shows the corresponding normalized oxygen distributions for the water. (X-ray experiments are insensitive to the hydrogens and only measure the oxygen spatial distribution). In our analysis, we have assumed that the X-ray scattering from the electrolyte is solely due to water and the scattering from Na^+ and F^- ions is negligible. This is justified for several reasons. First, neither ion significantly adsorbs on Ag(111) (ref. 21) and the interfacial ion concentration is thus small (≤ 0.2 in Fig. 4

as estimated from Gouy-Chapman theory¹). Second, the X-ray scattering strengths of Na^+ , F^- , and water are essentially identical, and the small ion concentrations will not significantly influence the scattering intensities. Third, measurements in 0.002 M NaF (not shown), where any ion adsorption would be significantly smaller, are the same as Fig. 3. These reasons provide confidence that Na^+ and F^- ions do not contribute significantly to the profiles in Fig. 4.

Figure 4 illustrates important features of the water structure. First, the water is ordered in layers extending about three molecular diameters from the electrode, and the layering is voltage dependent, suggesting that the interfacial electric field plays a role in the layering. Second, the spacing between the electrode and oxygens of the first layer is larger at -0.23 V than at +0.52 V, consistent with the idea of reorientation. However, the areal density in the first layer is enormous: 1.1 (-0.23 V) to 1.8 (+0.52 V) water molecules per Ag atom, whereas one expects ~ 0.8 molecules per Ag atom, based on the bulk water density. Thus, the first-layer water molecules are compressed by a factor of about two!

Because this huge areal density is surprising, it is important to assess the confidence we can place in our conclusions. We have checked many different models (including varying numbers of layers and constrained densities) and find that it is impossible to fit the data without the large areal densities. This analysis shows that the errors in these densities are $\sim \pm 15\%$. In addition, adequately fitting the data requires the first two water layers in Fig. 4a and the first three in Fig. 4b, while the peaks at larger z marginally improve the fits. An analysis similar to the Fourier

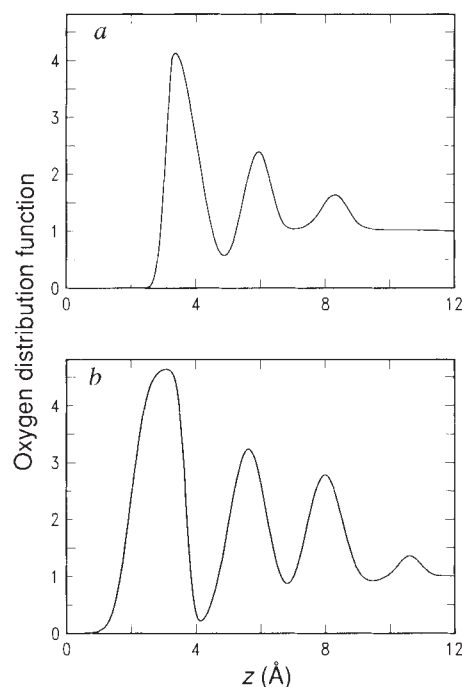


FIG. 4 a, b Best-fit models of the oxygen distributions for water near Ag(111) at -0.23 V of the p.z.c. (a) and +0.52 V of the p.z.c. (b). The oxygen distributions are normalized to their value far from the interface (that is, 3.34×10^{22} oxygens per cm^3), and hence approach unity for large z . Here z is distance above the top Ag(111) atomic plane. To model the oxygen distribution, we used gaussian functions for the second and higher layers and an error function (going from zero to unity) to represent the bulk water contribution. For the first layer, an asymmetric gaussian was necessary to fit the data for -0.23 V, whereas two opposing error functions separated by ~ 1 Å were needed for +0.52 V. The first-layer areal densities are 1.1 and 1.8 water molecules per Ag atom at -0.23 V and +0.52 V, respectively, with errors of $\sim \pm 15\%$.

difference synthesis²³ further supports our conclusions about the large areal densities and the extent of layering.

The layering we observe is consistent with that predicted by simulations and observed in other experiments, although the dependence on the magnitude of the applied voltage is unexpected. Furthermore, the electrode–oxygen spacings we observe provide direct experimental evidence for water reorientation from oxygen-up to oxygen-down as the electrode charge changes from negative to positive. Thus, our data support these aspects of the present picture of electrode–electrolyte interfaces. The large first-layer areal density has not, however, been predicted or observed in previous work. This large density implies that the first-layer water structure is determined by more complicated factors than steric packing, and that the hydrogen bonding in liquid water is not maintained in this layer. This will have consequences for interfacial dynamical and chemical properties (for example, viscosity, acidity). In addition, our results show that one cannot describe the first layer in terms of bulk-like water, and they are inconsistent with an ‘ice-like’ first layer, because the areal density of this is much smaller than our observations. Although more work is required before confident generalizations can be made to other electrodes, we can reasonably expect that our observations will qualitatively apply to aqueous–metal interfaces where there is no specific chemical interaction between water and the metal.

We offer one qualitative explanation of the large first-layer densities. At charged electrodes, the strong electric field ($\approx 10^7$ V cm⁻¹) orients the first-layer molecules through electrostatic interaction with the dipoles. Thus, water molecules in the bulk electrolyte feel an attractive force pulling them toward the surface, which increases the first-layer areal density. It should be noted that a similarly strong electric field exists around ions in solution, and that the relative areal densities we observe are comparable to the relative densities near solvated ions^{24,25}. Based on this explanation, we can reasonably speculate that the first-layer density has a minimum near the p.z.c. and increases as the magnitude of the applied voltage increases. Our results show that the present molecular models of water used in computer simulations and statistical models are inadequate for determining the properties of electrode–electrolyte interfaces. The current picture of how ions and molecules approach and adsorb at an interface will also be modified by our results, because in these processes the ion or molecule alters the interfacial water structure (including the dense first layer). Because the approach of a molecule to an interface is an important part of many heterogeneous chemical reactions (such as adsorption and electron transfer processes), our results change the picture of how this occurs at the atomic scale. □

22. Valette, G. & Hamelin, A. *J. electroanal. Chem.* **45**, 301–319 (1973).
23. Stout, G. H. & Jensen, L. H. *X-ray Structure Determination: A Practical Guide* (Wiley, New York, 1989).
24. Conway, B. E. *Chem. Soc. Rev.* 253–261 (1992).
25. Heinzinger, K. *Pure appl. Chem.* **57**, 1031–1042 (1985).

ACKNOWLEDGEMENTS. We thank J. Jordan-Sweet, B. Stephenson and R. Holaday for their assistance at NSLS and M. Philpott for useful discussions. This work was partially supported by the US Office of Naval Research and was performed at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the US Department of Energy.

Poor reproduction in forest passerines from decline of snail abundance on acidified soils

J. Graveland^{*†}, R. van der Wal[‡], J. H. van Balen^{*} & A. J. van Noordwijk^{*}

^{*} Netherlands Institute of Ecology, Centre for Terrestrial Ecology, PO Box 40, 6666 ZG Heteren, The Netherlands

[†] Present address: Institute for Forestry and Nature Research, PO Box 23, 6700 AA Wageningen, The Netherlands

[‡] Zoological Laboratory, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

ON poor, acidified soils in The Netherlands, an increasing number of great tits, *Parus major*, and other forest passerines, produce eggs with thin and porous shells¹. Here we show that the egg-shell defects, and the related high incidence of clutch desertion and empty nests, are caused by calcium deficiency, that snail shells are the main calcium source for the laying female, but that snails are scarce on poor soils. Similar laying irregularities in birds are reported from acidified regions elsewhere in Europe^{2–4}. We provide evidence that snails declined by a decrease in soil calcium on poor soils. Acid deposition is the main cause for decreasing calcium levels in such soils^{5–7}. To our knowledge, this is the first experimental evidence for calcium limitation in wild birds and it reveals a previously overlooked mechanism by which acidification affects higher trophic levels of the forest ecosystem.

In the Buunderkamp forest (20 km west of Arnhem in The Netherlands) study area, the proportion of great tits laying eggs with defective shells increased from 10% in 1983–84 to 40% in

TABLE 1 Differences between great tit eggs with normal and defective shells in Buunderkamp forest

Shell	Normal		Deviant		
	X ± s.d.	N	X ± s.d.	N	
Egg volume (mm ³)	1,346 ± 121	34	1,294 ± 178	22	NS*
Shell thickness (μm)	85.3 ± 5.6	26	77.1 ± 8.5	16	P < 0.01*
Egg calcium (mg)	30.6 ± 3.4	12	22.1 ± 4.0	8	P < 0.001*
Water content (%)	77.6 ± 0.9	15	57.6 ± 30.6	8	P < 0.01†
	% (n)		% (n)		
Hatched	95.3 (426)	447‡	3 (28)	31‡	P < 0.001§
Desiccated	0.5 (2)	447	55 (17)	31	P < 0.001§
Removed by parents	0.2 (1)	447	32 (10)	31	P < 0.001§
Sterile, dead embryo	4.0 (18)	447	3 (1)	31	NS§

Nests were checked daily during the laying period and new eggs were individually marked. Eggs were visually identified as normal or deviant on the day that they were laid; eggs with defective shells are easily recognized by their rough shell surface and deviant pigmentation. A sample of eggs was collected within a week of being laid to determine the shell thickness and other characteristics.

* Student's t-test.

† Mann–Whitney U-test.

§ χ^2 -test.

‡ In a total of 70 clutches (deviant shells present in 20 clutches).

|| Most eggs were removed after breakage of the shell.

Received 30 September 1993; accepted 21 January 1994.

1. Bockris, J. O'M., Conway, B. E. & Yeager, E. *Comprehensive Treatise of Electrochemistry* Vol. 1 (Plenum, New York, 1980).
2. Fawcett, W. R., Levine, S., deNobriga, R. M. & McDonald, A. C. *J. electroanal. Chem.* **111**, 163–180 (1980).
3. Lee, C. Y., McCammon, J. A. & Rossky, R. J. *J. chem. Phys.* **80**, 4448–4455 (1984).
4. Patey, G. N. & Torrie, G. M. *Chimica Scripta* **29A**, 39–47 (1989).
5. Schmickler, W. & Henderson, D. *Prog. Surf. Sci.* **22**, 323–420 (1986).
6. Price, D. & Halley, J. W. *Phys. Rev.* **B38**, 9357–9367 (1988).
7. Glosli, J. N. & Philpott, M. R. *J. chem. Phys.* **96**, 6962–6969 (1992).
8. Spohr, E. *J. phys. Chem.* **93**, 6171–6180 (1989).
9. Raghavan, K., Foster, K. & Berkowitz, M. *Chem. Phys. Lett.* **177**, 426–432 (1991).
10. Nagy, G. & Heinzinger, K. *J. electroanal. Chem.* **327**, 25–30 (1992).
11. Alosi, G., Foresti, M. L., Guidelli, R. & Barnes, P. J. *chem. Phys.* **91**, 5592–5596 (1989).
12. Porter, J. D. & Zinn, A. S. *J. phys. Chem.* **97**, 1190–1203 (1993).
13. Russell, A. E., Lin, A. S. & O'Grady, W. E. *J. chem. Soc., Faraday Trans.* **89**, 195–198 (1993).
14. Habib, M. A. & Bockris, J. O'M. *Langmuir* **2**, 388–392 (1986).
15. Wang, J., Ocko, A. J., Davenport, A. J. & Isaacs, H. S. *Phys. Rev.* **B46**, 10321–10338 (1992).
16. Pashley, R. M., Israelachvili, J. N. & Coll, J. *Interface Sci.* **101**, 511–523 (1984).
17. Toney, M. F. in *X-ray Methods in Corrosion and Interfacial Electrochemistry* (eds Davenport, A. J. & Gordon, J. G.) 1–24 (The Electrochemical Society, Pennington, 1992).
18. Toney, M. F. et al. *Phys. Rev.* **B45**, 9362–9374 (1992).
19. Feidenhans'l, R. *Surf. Sci. Rep.* **10**, 105–188 (1989).
20. Hamelin, A., Morin, S., Richer, J. & Lipkowski, J. *J. electroanal. Chem.* **272**, 241–252 (1989).
21. Valette, G. *J. electroanal. Chem.* **269**, 191–203 (1989).

TABLE 2 Comparison of the number of individual snails and of snail species in litter samples collected in oak forests in 1970–73 and 1992

	Hierden		Oldenaller		Voorst		Total*	
Rich soil	1973	1992	1973	1992	1973	1992	(NS)	NS
Individuals per m ² †	112 ± 88	120 ± 76 (NS‡)	35 ± 22	72 ± 57 (NS)	304 ± 178	206 ± 101	(NS)	NS
Species per sample	3.3 ± 1.2	3.4 ± 1.6 (NS)	1.1 ± 0.3	2.5 ± 1.8 (NS)	3.4 ± 1.6	4.8 ± 1.8	(0.05)	0.06
	De Nul		Zuurlander Esch		Total*			
Poor soil	1970	1992	1970	1992				
Individuals per m ²	13.7 ± 9.6	0.0 (<i>P</i> < 0.01)	44.9 ± 35.4	3.4 ± 5.0	(0.06) <i>P</i> < 0.001			
Species per sample	2.4 ± 0.9	0.0 (<i>P</i> < 0.01)	2.1 ± 1.0	0.5 ± 0.6	(0.06) <i>P</i> < 0.001			

Rich soil: three forests on nutrient-rich loamy sand soil. Poor soil: two forests on poor sand soil. The historical data were found by interviewing malacologists and searching through excursion reports. Sites were selected, where the sampling procedures and results were described accurately enough to permit resampling, and where the vegetation had changed only through natural succession. The data are based on ten samples of 25 × 25 cm² per year and per forest on rich soil, and on five (De Nul) or four (Zuurlander Esch) samples of 50 × 50 cm² on poor soil. *P* = 0.06 for the pooled data on rich soils indicates an increase in number of species, not a decrease.

* Significance for 1970–73/1992 comparison for all data combined.

† Calculated from numbers per sample (see description of methods).

‡ Mann–Whitney *U*-test.

1987–88¹. Buunderkamp is a mixed deciduous–coniferous forest on poor sandy soil and is representative of 80% of the Dutch forests (~2,500 km²). We investigated the cause and extent of the egg-shell defects and their effect on the breeding success.

The deviant shells are thin and porous and the eggs fail to hatch because of shell breakage and desiccation (Table 1). Females with defective shells deserted 48% (48 out of 100) of the clutches before hatching in Buunderkamp, and females with normal shells deserted 14% (34 out of 248, $\chi^2 = 46.52$, d.f. = 1, *P* < 0.001). About 10% of the nests remained empty, and some empty nests were incubated for up to three weeks. In four forests on poor soil (calcium content of the top 30 m of soil was 0.30 ± 0.08 g per kg dry matter), 40.8% of the females produced eggs with defective shells in 1991 (range among forests, 31–47%; *n* = 380 nests), and on rich soils (calcium content, 1.18 ± 0.55 g kg⁻¹; *n* = 3), 1.3% (range 0–4, *n* = 195).

The nature of the egg-shell defects and the relation with soil type suggested that the egg-shell defects were caused by calcium

deficiency. Poisoning by organo-chlorine compounds⁸ did not seem very likely, because raptors, which are more vulnerable to poisoning than passerines⁹, have fully recovered after the banning of DDT in The Netherlands¹⁰. The calcium contents of food types of great tits from rich soils were up to 50% higher than from poor soils (our unpublished results). However, we calculated that a diet of spiders (calcium content in 27 samples, 0.10 ± 0.09%, so higher than other food types) from rich soils would not cover more than 10% of the daily calcium demand of a laying female. There is little advance storage of calcium in the skeleton. Three females, collected after laying their first eggs, contained 206 mg (s.d. = 10) of calcium, only 7 mg more than three females collected on the last laying day. Eggs contain about 30 mg calcium and females produce nine eggs¹¹. Female great tits have to collect calcium-rich material, such as snail shells, during egg-laying, and calcium deficiency on poor soils could only result from a lack of calcium-rich material.

We tested the calcium-deficiency hypothesis by feeding experi-

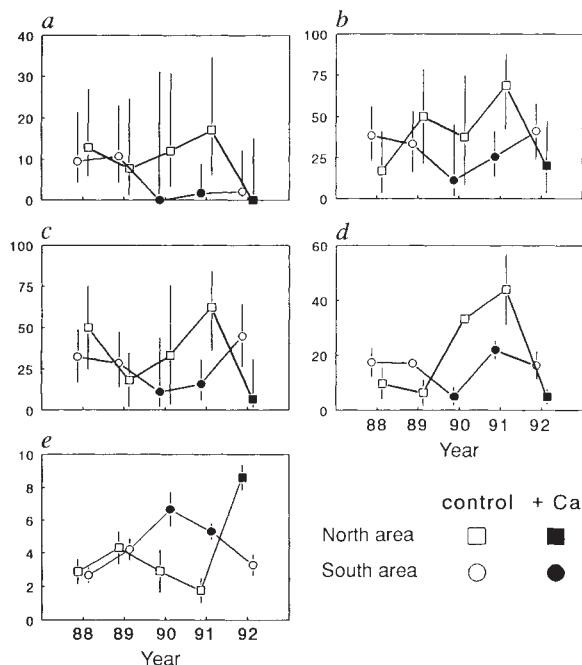


FIG. 1 Effect of the provision of Ca-rich material on the egg production and shell quality of great tits in Buunderkamp forest. filled symbols refer to calcium-fed birds, open symbols to unfed birds. a, Percentage of complete nests (with nest cup and lining) without eggs; b, percentage of clutches with at least one egg with a defective shell; c, percentage of clutches deserted before hatching (cases of predation and disturbance by observer excluded); d, percentage of unhatched eggs in clutches that were not deserted, but produced at least one hatchling; e, number of hatchlings per nesting attempt (including empty nests and deserted clutches). Vertical lines denote 95% confidence intervals (a–c) or s.e. (d, e). The high proportion of unhatched eggs in the calcium-fed birds in 1991 coincided with bad weather during the incubation period.

METHODS. Small feeders with five fragments of chicken eggshell and five fragments of snail shell (~100 mg each, 33.1% calcium in dry matter) were attached to all nestboxes in the southern (1990, 1991) or northern part (1992) of the forest around 10 March, about a month before the start of nest building. The feeders were refilled two-to-three times a week. Stealing by birds from adjacent territories precluded distribution of the feeders over the whole forest. Analysis: data were analysed with logistic regression (a–d) or discrete poisson regression (e, multiple regression not appropriate because of high proportion of nests (31%) without hatchlings). The independent variables were experiment, area and year (categorical, 1988–92). Significance was tested by dropping terms one at a time from a model with all the significant terms. The number of hatchlings (e) of the experimental birds was higher than of the control birds (change in deviance for dropping experiment from the full model 108.9, d.f. = 1, *F* = 27.3, *P* < 0.001) and differed between years (change in deviance 48.1, d.f. = 4, *F* = 3.02, *P* = 0.02; deviance null model 1,075.0, d.f. = 236, full model 921.1, d.f. = 231). The experiment contributed also significantly (*P* < 0.05) to the explained variance for the reproduction parameters a–d separately.

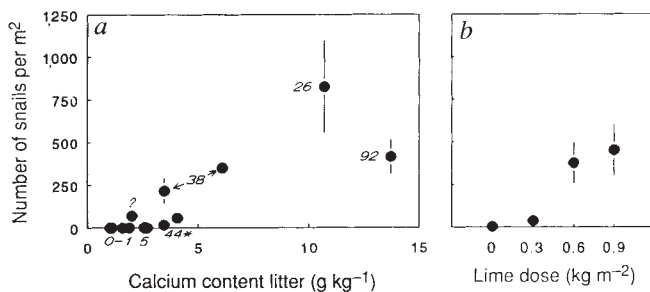


FIG. 2 Relationships between snail abundance, use of snails by great tits and the calcium content of the soil. *a*, Relation between the number of snails and the calcium content of the litter layer of oak forests ($r_s = 0.85$; $n = 14$ forests, $P < 0.001$). The samples were collected in June 1992 and consisted of the litter layer and the top 1 cm of the mineral soil of a 25×25 cm square. The points in the graph represent the means $\pm$ s.e. of 2–10 samples for calcium measurements (g kg^{-1} dry matter) and 5–10 samples for snail counts. The numbers at the data points denote the percentage of nests containing fragments of snail shells; 44* refers to two locations in a forest where great tits collected snail shells along paths paved with cockle shells. Snails were much more abundant there than at the sampling sites. *b*, Effect of liming on the number of snails in a 35-year-old oak forest on poor soil. Liming took place in 1988, with 0, 3, 6 or 9 tons of dolomitic lime (calcium magnesium carbonate) per hectare, in two series of 30×30 m plots. Date and sampling method as for *a*. The total number of snails was related to the lime dose (one-way ANOVA, $F_{[3,46]} = 8.85$, $P < 0.001$).

ments: great tits in Buunderkamp received snail shells and chicken egg shells in feeders attached to their nestboxes. There was a reduction among the fed birds in the number of empty nests, the incidence of egg-shell defects and clutch desertion, and the proportion of unhatched eggs (Fig. 1). We conclude that the egg-shell defects are caused by a lack of calcium-rich material on poor soils.

To determine the calcium sources used by great tits and which sources are lacking on poor soils, we examined the stomach content of five egg-laying females that had been collected in forests where egg-shell defects did not occur, and searched the nest material of nests on poor and rich soils for calcium-rich items. Nestlings receive calcium-rich material in addition to their normal food, caterpillars and spiders, which covers only 20% of the calcium demand for skeleton formation. The number of fragments of calcium-rich material found in nests can be used as an index of that fed to nestlings, because the number of fragments in nests and droppings of nestlings are correlated ($n = 160$ nests, chicken egg shells $r_s = 0.52$, $P < 0.001$; snail shells $r_s = 0.42$, $P < 0.001$). All females had snail shells in their stomach (62 mg, s.d. = 52), and no other calcium-rich material. Nests in four forests without egg-shell defects contained 4.4 (s.d. 4.8; 73 nests) snail-shell fragments (ranging among forests, 2.7–7.3). Nests from five forests on poor soils, where more than 30% of the females produced defective shells, contained 0.27 (s.d. 0.8; 420 nests) snail-shell fragments (0.0–0.7; Mann–Whitney U -test, $P < 0.001$). Chicken egg shells were the main calcium source on poor soils (3.1 ± 8.2 fragments per nest; more than snail shells, Wilcoxon signed rank test, $P < 0.001$). In forests not associated with egg-shell defects, chicken egg shells were less important (1.0 ± 2.3 fragments per nest). Chicken egg shells were collected outside the forest at farms and picnic sites. Birds nesting near such sites in Buunderkamp had more chicken egg shells in the nests, and fewer egg-shell defects (9%; 5 out of 55 clutches) than birds nesting elsewhere (35%, 37 out of 106; $\chi^2 = 12.52$,

$P < 0.001$). These findings imply that a lack of snail shells causes the eggshell defects on poor soils and that part of the birds try to compensate for the lack of snail shells by using anthropogenic calcium sources.

Natural snail abundance was correlated with the calcium content of the litter (Fig. 2a). A liming experiment on poor acidic soil resulted in snail densities comparable to those on rich soils (Fig. 2b). There were barely any snails in forests associated with high rates of egg-shell defects. The increase in egg-shell defects suggested that snails were more common in the past. We re-sampled old study sites and found that snail abundance and number of species had seriously declined in forests on poor soils but had remained the same in forests on rich sand (Table 2). A decline on poor soils, and no change on rich soils, were also found in Sweden¹². Snails require large amounts of calcium for reproduction and growth^{13–15} and obtain calcium, in addition to their normal food, by ingesting soil or rock, or by absorption through their skin¹⁶. Our results indicate that snail populations have declined as a result of a decrease in soil calcium. Acid deposition is the main cause for the decline of soil calcium on poor soils in The Netherlands and in acidified areas elsewhere^{5, 7, 17}. We conclude that acid deposition has caused calcium deficiency in tits through a decline in snail populations.

These results apply to other passerines as well. Females store little calcium in the skeleton and use calcium-rich material for shell formation^{18–22}. Snail shells are the most frequently used source of calcium^{18, 20, 22}. Until now, there was only circumstantial evidence for a relation between acid deposition and calcium shortage in birds, and the observed reduction in reproductive output was small^{23–25}. Our findings suggest that anthropogenic calcium sources can reduce the problem in densely populated countries such as The Netherlands, but calcium deficiency will cause severe reductions in songbird reproductive rates in more natural acidified habitats. □

Received 24 December 1993; accepted 7 February 1994.

1. Drent, P. J. & Woldendorp, J. W. *Nature* **339**, 431 (1989).
2. Carlsson, H. et al. *Orn. Svecica* **1**, 51–53 (1991).
3. Schmidt, K.-H. *J. Orn.* **131**, 172–174 (1990).
4. Winkel, W. & Huddle, H. *Vogelwarte* **35**, 341–350 (1990).
5. Hanson, D. W. et al. *Water Air Soil Pollut.* **18**, 227–239 (1990).
6. Tamm, C. O. & Hallbacken, L. *Ambio* **17**, 56–61 (1988).
7. Ulrich, B. Z. *Pflanzenernaehr. Boden.* **149**, 702–717 (1986).
8. Newton, I. & Bogan, J. *Nature* **249**, 582–583 (1974).
9. Newton, I. *Population Ecology of Raptors* (Poyser, Berkhamsted, UK, 1979).
10. Opdam, P. J. et al. *Ardea* **75**, 205–212 (1987).
11. Van Balen, J. H. *Ardea* **61**, 1–93 (1973).
12. Wärborn, I. *Ecography* **15**, 62–69 (1992).
13. Wärborn, I. *Oikos* **21**, 285–291 (1970).
14. Crowell, H. H. *Proc. Malac. Soc. Lond.* **40**, 491–503 (1973).
15. Ireland, M. P. *Comp. Biochem. Physiol. A* **98**, 111–116 (1991).
16. Gärdenfors, U. *Impact of Airborne Pollution on Terrestrial Invertebrates, with Particular Reference to Molluscs* (Report 3362, National Swedish Environment Protection Board, Stockholm, 1986).

17. Heij, G. J. & Schneider, T. (eds). *Acidification Research in The Netherlands* (Final Rep. of the Dutch Priority Programme on Acidification, Elsevier, Amsterdam, 1991).
18. Ankney, C. D. & Scott, D. M. *Auk* **97**, 684–696 (1980).
19. Schifferli, L. *Ornith. Beob.* **76**, 289–292 (1979).
20. Creutz, G. *Die Vogelwelt* **74**, 52–54 (1953).
21. Jones, P. J. *Ibis* **118**, 575–576 (1976).
22. Schifferli, L. *Ornithol. Beob.* **74**, 71–74 (1977).
23. Glooschenko, V. et al. *Water Air Soil Pollut.* **30**, 553–567 (1986).
24. Graveland, J. *Experientia* **46**, 962–970 (1990).
25. Ormerod, S. J. et al. *Environ. Pollut.* **55**, 107–121 (1988).

ACKNOWLEDGEMENTS. We thank C. M. Lessells, P. J. Drent, R. H. Drent, W. H. van der Putten and J. W. Woldendorp for comments on the manuscript; T. van Gijzen for help with collecting data; S. R. Troelstra and co-workers for chemical analyses; and the Forestry Service for allowing us to work in Buunderkamp forest. This study was supported by grants from the Department of Agriculture, Nature Management and Fisheries and the Department of Housing, Physical Planning and the Environment.

The first skull and other new discoveries of *Australopithecus afarensis* at Hadar, Ethiopia

William H. Kimbel*, Donald C. Johanson* & Yoel Rak*†

* Institute of Human Origins, 2453 Ridge Road, Berkeley, California 94709, USA

† Department of Anatomy, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

THE Hadar Formation in Ethiopia is a prolific source of Pliocene Hominidae attributed to the species *Australopithecus afarensis*¹. Since 1990, three seasons of field work have contributed 53 new specimens to the hominid inventory from Hadar, including the first fairly complete adult skull. Ranging from 3.0 to 3.4 million years in age (Fig. 1)²⁻⁴, the new specimens bear on key debates in hominid palaeontology, including the taxonomic implications of sample variation and the reconstruction of locomotor behaviour. They confirm the taxonomic unity of *A. afarensis* and constitute the largest body of evidence for about 0.9 million years of stasis in the earliest known hominid species.

The Hadar site yielded nearly 250 hominid fossils during the 1970s, including 40% of a female skeleton, A.L. 288-1 ('Lucy', ~3.18 million years), and the partial remains of at least 13 individuals from A.L. 333 (~3.20 Myr). Subsequent study of this collection and the sample from Laetoli, Tanzania, led to the recognition of the craniodentally primitive, bipedal species *Australopithecus afarensis*⁵. Morphological and metric variation in the *A. afarensis* hypodigm were initially explained as intra-specific, partly because of sexual dimorphism⁶. An alternative view attributed the differences to species-level taxonomic variation⁷. Debate was fuelled by the absence of a complete skull⁸⁻¹⁰. Similarly, although some interpreted the postcranial anatomy as indicating a complete adaptation to terrestrial bipedality^{11,12}, others saw evidence of a significant arboreal component in the locomotor repertoire of the species^{13,14}. The new Hadar discoveries address both of these issues.

The new skull A.L. 444-2, from the middle Kada Hadar (KH) Mb. (~3.0 Myr), is clearly that of an adult male (Fig. 2). It has large canines and, with a biasterionic breadth of 106 mm, it is the largest *Australopithecus* cranium on record. Although the maxillary postcanine teeth are larger than any of those in the limited Hadar + Laetoli sample from the 1970s, relative to canine size they are not unusual^{15,16}. The mandibular corpus is quite large, but its robusticity index falls below the Hadar mean (Table 1). The specimen's attribution to *A. afarensis* is warranted by its primitive constellation of mandibular, facial and calvarial characters relative to that of other *Australopithecus* species¹⁵⁻²⁰. Its morphology is consistent with that of Hadar specimens from A.L. 333 that comprise the major components of the composite

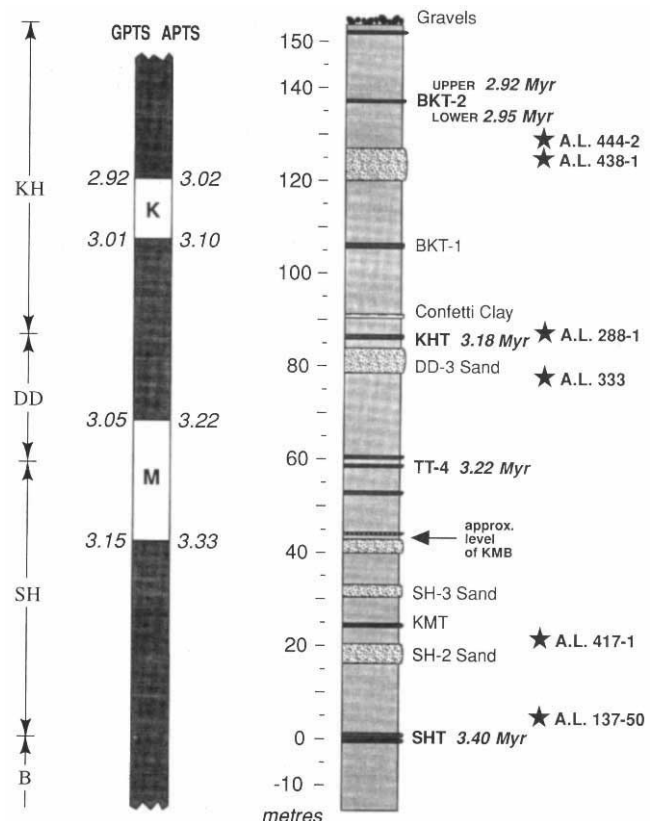


FIG. 1 Composite stratigraphic section of the Hadar Formation, based on revisions by R. C. Walter and J. L. Aronson through the 1993 field season, with positions of hominid specimens discussed in the text indicated. Single-crystal laser fusion (SCLF) ⁴⁰Ar/³⁹Ar ages for the SHT (3.40 ± 0.03 Myr), TT-4 (3.22 ± 0.01 Myr), KHT (3.18 ± 0.01 Myr) and BKT-2 (2.92–2.95 Myr) confirm that no Hadar hominid is older than 3.40 Myr, and establish ages of ~3.18 Myr for the A.L. 288-1 ('Lucy') skeleton and ~3.20 Myr for the A.L. 333 hominid assemblage^{2,3}. The Kadada Moumou basalt (KMB) does not occur in the type-area, but its relative stratigraphic position has been correlated into the type-section in the field. A whole-rock ⁴⁰Ar/³⁹Ar plateau age of 3.28 ± 0.04 Myr has been obtained for the KMB (ref. 4). New specimens from A.L. 438 and 444 represent the first Hadar hominid discoveries stratigraphically above A.L. 288-1 ('Lucy'), extending the temporal range of *A. afarensis* at Hadar from ~3.4 to ~3.0 Myr. At left are shown the stratigraphic ranges of Basal (B), Sidi Hakoma (SH), Denen Dora (DD) and Kada Hadar (KH) Members of the Hadar Formation. Also depicted is the composite magnetostratigraphic section with the global polarity time scale (GPTS) and the astronomical polarity time scale (APTS); the new ⁴⁰Ar/³⁹Ar and magnetostratigraphic data support the APTS calibration of the Mammoth (M) and Kaena (K) reversed subchrons^{3,4}.

reconstruction of a male *A. afarensis* skull based on the 1970s collection^{18,21,22}. The new Hadar skull refutes suggestions that the reconstruction incorporated facial and calvarial remains from two contemporaneous hominid species^{9,10}, and supports the interpretation of Hadar hominid cranial variation as intraspecific.

The frontal is a taxonomically important cranial region under-represented in the previous sample of *A. afarensis*. In A.L. 444-2 the supraorbital torus is vertically thick laterally and the very low squama has neither a chimpanzee-like supratrochlear sulcus nor a frontal trigone such as is found in 'robust' *Australopithecus* crania. The distance across the postorbital constriction in A.L. 444-2 is large compared to that of other *Australopithecus* species, both absolutely (77 mm) and relative to facial breadth (frontobiorbital breadth index ~80%)¹⁸. These features are shared with the ~3.9 Myr Belohdelie frontal BEL-VP-1/1 (ref. 23), confirming the previously tentative attribution of the latter specimen to *A. afarensis*²⁴.

TABLE 1 Mandible corpus measurements

	Breadth (mm)	Height (mm)	Robusticity (%)
A.L. 417-1a (left)	18.0	36.0	50.0
A.L. 444-2b (right)	22.0	43.2	50.9
1970s Hadar sample	$\bar{x}$ = 18.5 n = 12 σ = 2.1 range = 15.6–22.4	$\bar{x}$ = 33.6 n = 10 σ = 4.1 range = 28.0–40.5	$\bar{x}$ = 55.1 n = 10 σ = 6.4 range = 45.5–63.8

Mandible corpus measurements of A.L. 417-1a, compared with those of A.L. 444-2b and the 1970s Hadar sample²⁶. All measurements recorded here are for P4/M1 position, since measurements at M1 on A.L. 444-2b are not possible without further reconstruction of the corpus.

Maxilla A.L. 417-1d from the middle Sidi Hakoma (SH) Mb. (~3.25 Myr), associated with a partial mandible and cranial base fragments, constitutes the first relatively intact face of an *A. afarensis* female (Fig. 3a-d). It features small canines and less prognathism than A.L. 444-2, consistent with the great ape model of intraspecific sexual dimorphism^{18,25}, and otherwise demonstrates the same morphological pattern as do larger *A. afarensis* faces¹⁷. The mandible conforms to the *A. afarensis* plan²⁶. Its low corpus robusticity index is identical to that of the A.L. 444-2 male mandible (Table 1). The P₃ lacks a lingual cusp, as in other *A. afarensis* females (A.L. 128-23, 288-1i) and some males (A.L. 277-1), but another new female jaw, A.L. 315-22, has a bicuspid P₃, confirming that P₃ polymorphism cuts across the size range of Hadar mandibles.

The A.L. 438-1 partial upper limb skeleton from middle KH Mb. (~3.0 Myr) includes the most complete *Australopithecus* ulna known (Fig. 3e, f); based on the associated mandible (corpus height at P₄/M₁ = 42.5 mm) and frontal fragment, it is almost certainly a male. Maximum ulnar length (excluding styloid process) of 268 mm is 22% greater than our length estimate for the A.L. 288-1 female ulna (220 mm). The size disparity between these two Hadar ulnae is equalled by proximal ulnae A.L. 333x-5 and 333w-36. With anteroposterior shaft curvature

visibly less than in chimpanzees, a tall olecranon process, and an anteriorly oriented trochlear notch, the A.L. 438-1 ulna diverges from the modern African ape condition and matches the human morphology of other Hadar hominid ulnae^{27,28}.

The large, presumably male, humerus A.L. 137-50 from the lower SH Mb. (Fig. 3e, f) is a remarkably close match for the recently reported Maka humerus MAK-VP-1/3, with which it is penecontemporaneous (~3.4 Myr)²³, as well as for Hadar proximal humerus A.L. 333-107 (~3.2 Myr)²⁸. The new Hadar specimen has a retroflexed shaft and appears to be very robust relative to length, with thick cortical bone, pronounced deltoid and humeral adductor insertion sites, and a strong lateral supracondylar crest.

Our preliminary estimate of humeral length for A.L. 137-50 is 295 mm, ~24% greater than that of A.L. 288-1 (238 mm, ref. 27). Using the A.L. 438-1 ulna as a counterpart to A.L. 137-50 yields an ulna/humerus length index of ~91%. The index for the A.L. 288-1 female is ~92.5%. If it is true that in *A. afarensis* humerus length is roughly that expected in a modern human of comparable body size²⁹, then these figures indicate relatively long ulnae (and, presumably, forearms) for the Hadar hominids, distinctly closer to the relative ulnar length of chimpanzees ($\bar{x}$ = 95%, σ = 3%, r = 88–101%, n = 20) than to that of modern

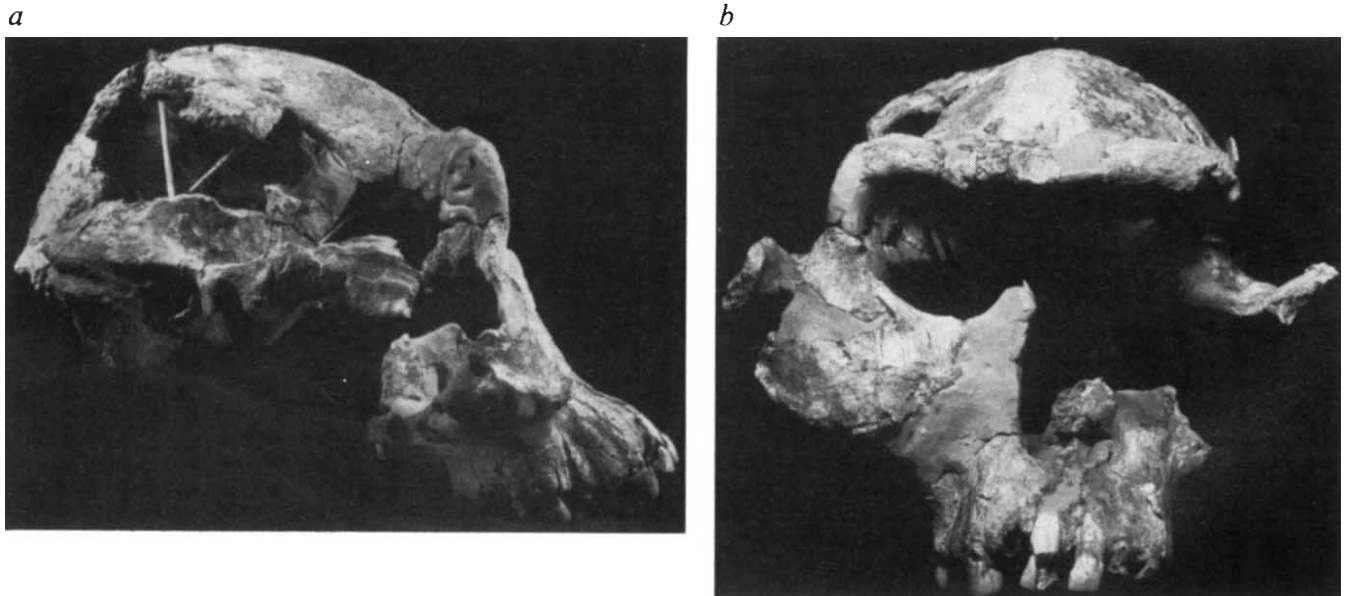


FIG. 2 Lateral (a) and anterior (b) views of Hadar skull A.L. 444-2, preliminary reconstruction, 47.5% natural size. Thirteen major pieces and many small fragments were clustered together in a small gully on a steep surface of Kada Hadar Member silts. Excavation failed to reveal additional hominid remains, but *in situ* mammal bone fragments with patina and preservation identical to that of the hominid were recovered upslope from the skull in a stratigraphic level 10.5 m below marker tuff BKT-2 (Lower). We believe this level to be the source horizon of the skull (Fig. 1). It comprises the right mandible and symphysis (not shown) with fragmentary left and right incisors, partial right C and damaged P₄-M₁, the maxilla with right I¹, C, P⁴-M³ and left I¹, C, P³-M³, the right zygomatic, most of the nasal bones (not shown), most of the frontal with attached anteromedial segments of the parietals, the left parietal, the occipital with attached posterior right parietal, and both temporals. The anterior cranial base is missing, as are most of both temporal squames. Although there are no contacts between the frontal, the maxilla and the zygomatic, there is sufficient preserved to provide a preliminary idea of their positions relative to one another and to the calvaria. Post-depositional deformation has compressed the palate along the midline, uplifted the right side of the frontoparietal fragment and supraorbital, twisted the left and elevated the right zygomatic arch, and pushed the nuchal plane of the occipital into the cranial cavity. The *A. afarensis* character complex present in A.L. 444-2 comprises a higher

number of hominid symplesiomorphies than in any other species of the genus (including those attributed by some to *Paranthropus*). These primitive characters include strong prognathism relative to both the calvaria and the P⁴/M¹ take-off of the zygomatic process of the maxilla, marked projection of sagittally and transversely convex premaxilla anterior to bicanine plane, horizontally inclined nasoalveolar clivus, procumbent maxillary incisors, extensive intranasal component of premaxilla with step-down to nasal cavity floor, sharp lateral nasal aperture margins, canine fossae, curved zygomaticoalveolar crests, anteriorly flat palate, I²/C + C/P₃ diastemata, posteriorly convergent temporal lines with low, probably bifid, posterior sagittal crest, compound temporal/nuchal crest, asterionic notch sutural pattern, massive mastoid processes inflected beneath the cranial base with tips independent of occipitomastoid crests, shallow mandibular fossae, tubular tympanic plates that sit completely posterior to large postglenoid processes, weak petrous crest of tympanics, and coexistent transverse-sigmoid and occipital-marginal venous drainage systems (the polarity of this last character is indeterminate). Estimation of cranial capacity should be possible, but awaits final reconstruction of the calvaria. Measurements of the better-preserved maxillary teeth are (in mm, buccolingual breadth × mesiodistal length, the latter corrected for interproximal wear): Right C: 11.5 × 10.4; Left P₄: 14.5 × 10.6, LM1: 15.0 × 13.7, LM2: 15.8 × 14.2, LM3: 16.2 × 14.9.

FIG. 3 Lateral (a) and anterior (b) views of A.L. 417-1d maxilla; lateral (c) and occlusal (d) views of A.L. 417-1a mandible, 50% natural size. Diagnostic *A. afarensis* morphology combined in the specimen is primitive relative to that of other *Australopithecus* species and includes: mandible corpus hollowed laterally, mental foramen positioned below midcorpus and opens anterosuperiorly, extramolar sulcus high and narrow, postcanine tooth row slightly concave buccally, low corpus robusticity, flat, vertical infraorbital bone plate, sagittally and transversely convex premaxilla that protrudes in front of the bicanine plane, and sharp lateral nasal aperture margin. Measurements of the dentition are (recorded as described in Fig. 2 legend): Maxilla, RC: 9.9×9.4 , RP3: 11.7×8.3 , RP4: 12.2×8.9 , RM1: 13.4×12.3 , RM2: 14.7×13.2 , RM3: 14.8×13.0 ; Mandible, LP3: 10.8×8.9 , LP4: 11.2×8.6 , LM1: 11.6×12.4 , LM2: 13.1×13.0 ; RM3: 13.3×14.9 . e, Anterior view of right humerus A.L. 137-50; f, lateral view of left ulna A.L. 438-1a, both 50% natural size. The A.L. 438 ulna is associated with a partial mandible, a frontal fragment, three metacarpals, as well as shaft portions of the right ulna, left and right radii and right humerus. Two fragments of the complete ulna were recovered *in situ* in a KH Member sand unit, 12 m below marker tuff BKT-2 (Lower) (Fig. 1).



humans ($\bar{x} = 80\%$, $\sigma = 3\%$, $r = 74-88\%$, $n = 20$). The combination of a relatively short, but robust, humerus and a long forearm is unlikely to resolve the debate about locomotion in *A. afarensis*—which has been concerned as much with incompatible evolutionary models for the interpretation of functional morphology as with divergent interpretations of the fossils themselves^{11,12,30}.

The new Hadar fossils support the taxonomic unity of *A. afarensis*. A single hominid species is now well documented in all three fossiliferous members of the Hadar Formation, spanning 0.4 Myr. The new discoveries for the first time extend the *A. afarensis* stratigraphic range well up into the Kada Hadar Member, by which time in the Hadar Fm. a variety of other macro-mammal clades had gone through taxonomic turnover³¹. Furthermore, with ages of ~ 3.0 Myr for the A.L. 444-2 skull

and ~ 3.9 Myr for the Belohdelie frontal, the 0.9 Myr temporal range of *A. afarensis* is bracketed by specimens of high diagnostic value. The Hadar sample forms nearly all of the post-3.4 Myr portion of the *A. afarensis* hypodigm and, at close to 300 specimens, it is the most persuasive body of evidence for prolonged stasis within the oldest known hominid species^{6,23}. □

Received 10 January; accepted 22 February 1994.

- Johanson, D. C., Taieb, M. & Coppens, Y. *Am. J. phys. Anthrop.* **57**, 373–402 (1982).
- Walter, R. C. & Aronson, J. L. *J. hum. Evol.* **25**, 229–240 (1993).
- Walter, R. C. *Geology* **22**, 6–10 (1994).
- Renne, P., Walter, R., Verosub, K., Sweitzer, M. & Aronson, J. *Geophys. Res. Lett.* **20**, 1067–1070 (1993).
- Johanson, D. C., White, T. D. & Coppens, Y. *Kirtlandia* **28**, 1–14 (1978).
- Johanson, D. C. & White, T. D. *Science* **203**, 321–329 (1979).
- Olson, T. R. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 102–119 (Liss, New York, 1985).
- Leakey, R. E. & Walker, A. C. *Science* **207**, 1103 (1980).
- Shipman, P. *Discover* **7**, 87–93 (1986).
- McKee, J. K. *Am. J. phys. Anthrop.* **80**, 1–9 (1989).
- Lovejoy, C. O. *Sci. Am.* **256**, 118–125 (1988).
- Latimer, B. M. in *Origine(s) de la Bipédie chez les Hominidés* (eds Coppens, Y. & Senut, B.) 169–176 (CNRS, Paris, 1991).
- Susman, R. L., Stern, J. T. & Jungers, W. L. *Folia primatol.* **43**, 113–156 (1984).
- Stern, J. T. & Susman, R. L. *Am. J. phys. Anthrop.* **60**, 279–318 (1983).
- White, T. D., Johanson, D. C. & Kimbel, W. H. S. *Afr. J. Sci.* **77**, 445–470 (1981).
- Kimbel, W. H., White, T. D. & Johanson, D. C. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 122–137 (Liss, New York, 1985).
- Rak, Y. *The Australopithecine Face* 1–169 (Academic, New York, 1983).
- Kimbel, W. H., White, T. D. & Johanson, D. C. *Am. J. phys. Anthrop.* **64**, 337–388 (1984).
- Kimbel, W. H. & Rak, Y. *Am. J. phys. Anthrop.* **66**, 31–54 (1985).
- Ward, S. C. & Kimbel, W. H. *Am. J. phys. Anthrop.* **61**, 157–171 (1983).

- Kimbel, W. H., Johanson, D. C. & Coppens, Y. *Am. J. phys. Anthrop.* **57**, 453–499 (1982).
- Kimbel, W. H. & White, T. D. *J. hum. Evol.* **17**, 545–550 (1988).
- White, T. D. et al. *Nature* **366**, 261–265 (1993).
- Asfaw, B. *J. hum. Evol.* **16**, 611–624 (1987).
- Wood, B. A., Yu, L. & Willoughby, C. *J. Anat.* **174**, 185–205 (1991).
- White, T. D. & Johanson, D. C. *Am. J. phys. Anthrop.* **57**, 501–544 (1982).
- Johanson, D. C. et al. *Am. J. phys. Anthrop.* **57**, 403–451 (1982).
- Lovejoy, C. O., Johanson, D. C. & Coppens, Y. *Am. J. phys. Anthrop.* **57**, 637–650 (1982).
- Jungers, W. L. & Stern, J. T. *J. hum. Evol.* **12**, 673–684 (1983).
- Stern, J. T. & Susman, R. L. in *Origine(s) de la Bipédie chez les Hominidés* (eds Coppens, Y. & Senut, B.) 99–111 (CNRS, Paris, 1991).
- White, T. D., Moore, R. V. & Suwa, G. *J. vert. Paleont.* **4**, 575–583 (1984).

ACKNOWLEDGEMENTS. We thank the Center for Research and Conservation of Cultural Heritage (Ethiopian Ministry of Culture and Sports Affairs) and the National Museum of Ethiopia for permission to conduct field work and for logistical support. Field work was funded by grants from the National Science Foundation, the National Geographic Society, the L. S. B. Leakey Foundation and Lufthansa Airlines. We thank R. Walter, chief geologist and project co-leader, J. Aronson, R. Bernor, M. Black, G. Eck, J. Harris, E. Hovers, N. Kahn, P. Renne, S. Semaw, C. Vondra and T. Yemane for scientific contributions to the project, T. White for the discovery of A.L. 137-50 during his visit to Hadar in 1990, B. Latimer and L. Jellema for human and chimpanzee data from the Hamann-Todd Osteological Collection, Cleveland Museum of Natural History and our Ethiopian colleagues T. Asework, A. Asfaw, Z. Assefa, H. Bolku, M. Fesseha, T. Hagos, M. Kassa, M. Tesfaye, S. Teshome, T. Wodajo and the Afar people of Eloha village without whose assistance our field work would not have been possible.

Matrix correlation tests support a single origin for modern humans

Diane M. Waddle*

Doctoral Program in Anthropological Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

THE debate over human origins has focused on two competing theories. The single African origin model holds that anatomically modern *Homo sapiens* evolved in Africa 100,000–200,000 years ago. Members of this population migrated out of Africa, replacing archaic human groups through Asia and Europe, with racial differentiation occurring within the past 100,000 years^{1–3}. The alternative regional continuity model proposes continuous evolution over the past million years, with racial variation developing early, and similar modern human traits developing in all regions as the result of worldwide gene flow^{4–6}. The persistence of specific morphological features within regions over the past million years supports regional continuity, whereas the identification of anatomically modern fossil specimens from Africa and the Levant 50–60,000 years before they are found elsewhere, provides support for a single origin. I give here the first quantitative test of the fossil evidence for each of these models. Results support a single African and/or Levantine origin for modern humans.

Matrix correlation methods are commonly used to evaluate the correspondence between two types of measured distances—for example, genetic, morphological, linguistic and geographical⁷. The present study uses matrix correlation methods to compare a set of morphological distances with a set of 'hypothetical' distances (the distances that are predicted to exist under a particular model of modern human origins) to identify the model that best explains observed variation among fossil crania (see ref. 8 for another example of this use).

A sample of 83 fossil crania was divided into 12 samples, or operational taxonomic units (OTUs)⁹ (Fig. 1a). The morphological distance between each pair of samples was calculated and the values arranged in a morphological distance matrix (Fig. 1b).

Connection schemes and design matrices (described below and in Fig. 2 legend) are used to represent specific models of modern human origins as numerical matrices of the same size as the matrix in Fig. 1b. Connection schemes describe the strength of links between populations that are predicted to exist under a particular model. Figure 2a shows such a scheme for a regional continuity model. In this model, morphological similarities among populations result from the opposing forces of evolutionary descent within regions and gene flow across regions. A single African origin model is shown in Fig. 2b. This model holds that all anatomically modern humans are descended from a single source population that lived in Africa up to 125,000 yr ago (OTU 9). Hypothetical distance values have been assigned to each of the links in these connection schemes and the design matrices that are derived from them are shown in Fig. 2c and d. Note that OTUs are linked and weights are assigned to the links for external reasons, that is, based on published descriptions of models of modern human origins, not with the intention of reproducing the values in the morphological distance matrix.

The correlation between design matrices representing models of modern human origins (such as those in Fig. 2c and d) and the morphological distance matrix of Fig. 1b is evaluated by the Mantel *r* statistic, which ranges from –1.0 for a perfect negative, to 1.0 for a perfect positive correlation between two matrices^{7,10}. Significance of the correlation is determined by a permutation test; the rows and columns of one matrix are permuted and the

Mantel statistic calculated 499 times, creating a distribution that is used to evaluate the significance of the observed correlation.

Results of matrix correlations for various models of modern human origins are given in Table 1. Tests of the models in Fig. 2 (models 1 and 2) show that the single African origin model provides a better explanation for the cranial variation described in the morphological distance matrix of Fig. 1b. Because OTU 9 comprises a morphologically diverse set of fossils, some of which

a	WEST EUROPE				EAST EUROPE		SOUTHWEST ASIA		AFRICA	
	(OTU 1)				(OTU 2)		(OTU 3)		(OTU 4)	
8	CHANCELADE 1 CRO MAGNON 1, 2 GOIGUES CAVE 1 OBERKASSEL 1, 2 ST GERMAINE LA RIVIERE 4 TEVIEC 1, 8, 9, 16				BRNO 2 DOLNI VESTONICE 3 MLADEC 1, 2 STETTEN 1, 2		EL WAD 4, 7, 8, 33 ERQ ELAHMAR 2 KEBARA H4 NAHAL EN GEV 1 SHUKBAH 12		AFALOU 3, 9, 29, 30, 34 ELANDS BAY - ALBANY ELEMENTETA B, D FISH HOEK 1 GAMBLES CAVE 4 MATJES RIVER X MUMBWA IV-3 NAIVASHA NVI, NV2, N1 OAKHURST 5 WADI HALFA 14, 20, 25, 30, 32, 34, 36, 37	
32	(OTU 5)				(OTU 6)		(OTU 7)	(OTU 8)	(OTU 9)	
kyr	GIBRALTAR 1 GUATTARI 1 LA CHAPELLE 1 LA FERRASSIE 1 LA QUINA H5 SACCOPASTORE 1, 2 SPY 1, 2				KRAPINA C3		AMUD 1 SHANIDAR 1 TABUN C1	QAFZEH 6, 9 SKHUL 4, 5, 9	BORDER CAVE 1 FLORISBAD 1 JEBEL IRHOUD 1, 2	
125	(OTU 10)				(OTU 11)		(OTU 12)			
600	ARAGO 21/47 PETRALONA 1 STEINHEIM 1				ZUTTIYEH 1		BODO 1 BROKEN HILL 1 ELIYE SPRINGS EVASI 1 LAETOLI 18 NDUTU 1 OMO 2 SALDANHA 1			

b	OTU 1	OTU 2	OTU 3	OTU 4	OTU 5	OTU 6	OTU 7	OTU 8	OTU 9	OTU 10	OTU 11	OTU 12
	0.00											
	0.88	0.00										
	0.92	0.96	0.00									
	0.81	0.89	0.89	0.00								
	1.53	1.53	1.70	1.48	0.00							
	1.36	1.44	1.71	1.57	1.14	0.00						
	1.45	1.58	1.62	1.54	0.96	1.00	0.00					
	1.03	1.14	1.16	1.02	1.26	1.14	1.32	0.00				
	1.50	1.67	1.64	1.60	1.63	1.51	1.60	1.43	0.00			
	1.61	1.63	1.70	1.53	1.39	1.27	1.55	1.32	1.62	0.00		
	1.31	1.45	1.40	1.16	1.13	1.21	1.22	0.96	1.69	1.34	0.00	
	1.62	1.72	1.81	1.56	1.31	1.69	1.55	1.34	1.58	1.49	1.66	0.00
	OTU 1	2	3	4	5	6	7	8	9	10	11	12

FIG. 1 a, Specimens used in this study were divided into 12 OTUs by geographical location of the fossil site and chronological age of the specimen reported in literature sources. Matrix correlations are reported in Table 1 for a second set of 12 OTUs (not shown here) which differs from this arrangement in the transfer of Laetoli 18 from OTU 12 to OTU 9, and Jebel Irhoud 1 and 2 from OTU 9 to OTU 12. Chronological dates reported for these specimens permit their inclusion in OTU 9 or OTU 12 (refs 13, 15, 20, 21). Due to lack of precision in the dating of fossil sites, other specimens might also be allocated to alternative OTUs (for example, Florisbad²²). However, small changes in OTU composition, and/or in connection scheme weights, have little effect on the results of specific matrix comparisons (Table 1; and ref. 17), and no effect on the overall results of this study. Although the expression of models of modern human origins as design matrices necessarily lacks fine detail, the robustness of the results to changes in the analysis strengthen the conclusions reached in this letter. b, Matrix expressing morphological distance between all pairs of OTUs in a. Cranial data used to calculate this matrix consist of 74 metric traits (for example, cranial lengths and breadths, as described by Howells²³), 52 discrete traits coded as present or absent (including defining features of *Homo erectus*, Neanderthals, anatomically modern humans), and 39 angles and indices calculated from linear measurements¹⁷. Variable means were calculated for all OTUs; means were standardized across OTUs to yield a grand mean of 0.0 and a standard deviation of 1.0. Missing values were entered as 'no record' and were not included in the calculation of mean variable values. Morphological distances were calculated between OTU pairs using the formula for average taxonomic distance⁹. Missing data for an OTU are not included in the calculations; pairwise taxonomic distance is calculated using the set of variables that are available for both OTUs of a pair.

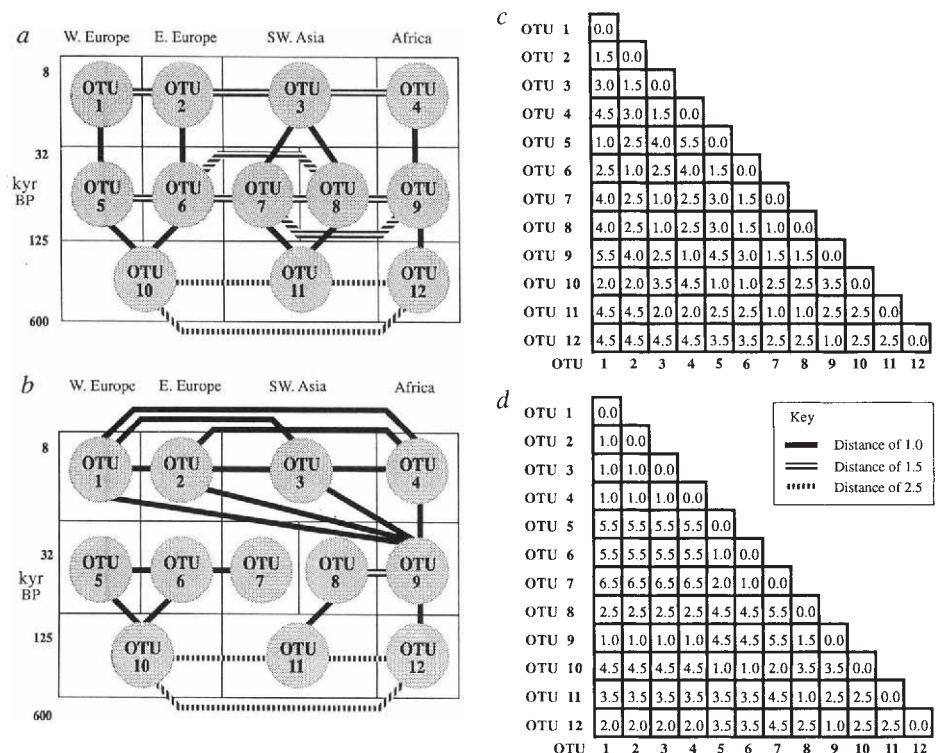
*Present address: Department of Biological Anthropology and Anatomy, PO Box 3170, Duke University Medical Center, Durham, North Carolina 27710, USA.

TABLE 1 Matrix correlations for models of modern human origins

	Model	<i>r</i>	<i>P</i>	<i>t</i> _{approx.}
1	Regional continuity (Fig. 2a)	0.257 (0.248)	0.060 (0.050)	1.924 (1.916)
2	Single African origin (Fig. 2b)	0.407 (0.533)	0.010 (0.002)	3.129 (4.186)
	Ancestral population represented by:			
2.1	Border Cave 1	0.469	0.002	3.624
2.2	Florisbad	0.335	0.010	2.574
2.3	Jebel Irhoud 1 and 2	0.382	0.012	2.943
3	Regional continuity: gene flow and continuity assigned equal weight	0.334 (0.318)	0.030 (0.008)	2.401 (2.396)
4	Regional continuity: strong gene flow between 125 kyr BP and 8 kyr BP	0.451 (0.442)	0.002 (0.002)	3.124 (3.261)
5	Regional continuity: strong gene flow between 125 and 32 kyr BP, stronger gene flow between 32 and 8 kyr BP	0.493 (0.477)	0.002 (0.002)	3.266 (3.424)
6	Single SW Asian origin (Fig. 3)	0.577 (0.656)	0.002 (0.002)	4.439 (5.157)
	Ancestral population represented by:			
6.1	Qafzeh 6	0.550	0.002	4.282
6.2	Qafzeh 9	0.499	0.002	3.878
6.3	Skhul 4	0.438	0.006	3.439
6.4	Skhul 5	0.544	0.002	4.243
6.5	Skhul 9	0.434	0.006	3.375
7	African + SW Asian origin	0.419 (0.546)	0.004 (0.002)	3.082 (4.178)

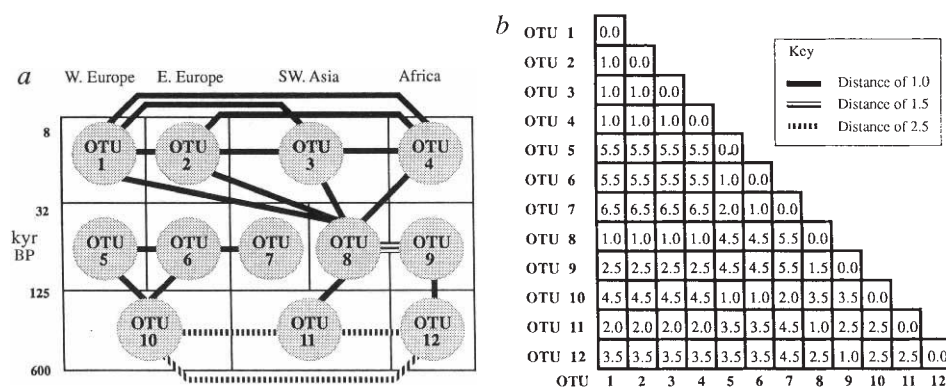
Correlations between models depicted as design matrices (see, for example, Figs 2c, d and 3b) and the morphological distance matrix of Fig. 1b. These results were not changed by the removal of certain individual specimens with questionable provenance, for example, Border Cave^{18,19}. Results of matrix correlations using the revised OTUs described in Fig. 1a legend are given in parentheses. Reported values include the Mantel correlation (*r*), probability (*P*) and approximate *t* value (*t*_{approx.})¹⁰.

FIG. 2 Connection schemes and design matrices for two different models of modern human origins. Distances between OTUs can be calculated using the key on the right of the figure. a, Connection scheme among the 12 OTUs under a regional continuity model. In this model modern humans (OTUs 1–4) are descended from fossil populations of the same regions. The closest similarities between fossil groups are predicted to exist within regions, over time periods. Gene flow across all regions, within each time period, is predicted to be two-thirds as strong as evolutionary descent in contributing to the measure of similarity between populations. In this scheme and the single African origin connection scheme (b), the specimens of the earliest time period considered, commonly referred to 'archaic' *Homo sapiens*, are connected by a relatively small distance value to indicate their predicted recent common ancestry. b, Connection scheme among the 12 OTUs for a single African origin model. All modern humans are derived from an African source population represented by OTU 9. European and Levantine Neanderthals (OTUs 5, 6 and 7) are derived from a European archaic population and are hypothesized to have become extinct without contributing to the modern human gene pool through interbreeding^{24,25}. The Skhul and Qafzeh hominids (represented by OTU 8), which have been described as 'anatomically modern', are predicted to have interbred with members of the African population,



but make no contribution to the modern human gene pool. c, d, Design matrices for the connection schemes of a and b, respectively, giving design matrix distances among all pairs of OTUs. Distances given in these matrices are calculated by adding together the values along the shortest path between any given pair of OTUs using the key.

FIG. 3 a, Connection scheme among the 12 OTUs for a single southwest Asian origin model. All modern humans are derived from a source population represented by the hominid fossils from the Israeli sites of Skhul and Qafzeh. As in the single African origin model, Neanderthals (OTUs 5, 6 and 7) are hypothesized to have become extinct without contributing to the modern human gene pool. The African hominids living from 32 to 125 kyr BP (OTU 9) are predicted to have interbred with members of the Levantine population. b, Design matrix for the connection scheme of a. Distances given in this matrix are calculated by adding together the values along the shortest path between any given pair of OTUs using the key.



are poorly dated and/or poorly provenanced, the single African origin model was tested against morphological distance matrices calculated using individual specimens to represent the ancestral African population (Table 1; models 2.1–2.3). The pattern of high and low correlations is consistent with the more modern morphology of Border Cave over the other specimens. The addition of well-preserved, accurately dated specimens is needed to test this model further.

Exact levels of predicted Pleistocene gene flow are not specified in published discussions of regional continuity. Therefore, to avoid bias against the regional continuity model resulting from the weight assigned to gene flow in Fig. 2a, several continuity models were tested in which the relative weights of gene flow and continuity were varied. Correlations between the regional continuity model and the set of craniometric distances are improved considerably by increasing the strength of gene flow relative to evolutionary continuity. When gene flow and continuity are weighted equally, the correlation increases (model 3). When the strength of gene flow is double that of continuity (model 4), and when the level of gene flow in the most recent time period is four times the weight assigned to continuity (model 5), the correlation increases further, although it remains less than that achieved using the alternative set of OTUs (Fig. 1a; model 2). These results support the gradistic notion that coterminous taxa tend to be similar in morphology. They do not support the idea of evolutionary continuity within regions as indicated by the persistence of specific cranial features.

Given that some of the earliest anatomically modern *Homo sapiens* fossils have been found at the Israeli sites of Skhul and Qafzeh^{11–16}, southwest Asia is a possible site of origin for the first modern humans. The hypothesis that modern human morphology first developed in the Levant was tested using a single southwest Asian origin model (Fig. 3a). The correlation between

the design matrix derived from this model (Fig. 3b; model 6) and the morphological distance matrix is higher than that produced under the single African origin model (model 2). Using the same methods as for the African ancestral population, this model was tested against morphological distance matrices that used individual specimens to represent the ancestral population, OTU 8 (models 6.1–6.5). All the correlations are highly significant. The lowest correlations are produced using specimens with the most missing data.

Finally, a single-origin model was tested in which the ancestral population was represented by a combined set of specimens from Africa and the Levant. This model was constructed by modifying the connection scheme of Fig. 2c so that the distance from OTU 8 to OTU 9 is 0.0. This model (model 7) is also highly correlated with the morphological distance matrix, but does not perform as well as the single southwest Asian origin model.

The quantitative analyses presented here support a single origin for modern humans as opposed to continuous long-term evolution within regions. Models that hypothesize worldwide gene flow in opposition to evolutionary continuity explain observed fossil variation only if continuity is a relatively weak force compared with gene flow. Results of tests to be published elsewhere confirm that these results hold when matrix correlation tests are undertaken between 83×83 matrices of individual specimens, rather than OTUs¹⁷. The single-origin model is also supported by tests against morphological distance matrices calculated on discrete traits alone, and on size-free traits alone¹⁷. The results presented here cannot resolve the issue of the site of origin. This study clearly refutes evolutionary continuity in Europe, and for the Neanderthals of southwest Asia. Studies are underway to use these same matrix correlation methods in tests of regional continuity and single-origin models incorporating the east Asian fossil record. □

Received 26 July; accepted 22 December 1993.

- Stringer, C. B. & Andrews, P. *Science* **239**, 1263–1264 (1988).
- Stringer, C. B. & Andrews, P. *Science* **241**, 773–774 (1988).
- Delson, E. *Nature* **332**, 206 (1988).
- Wolpoff, M. H. in *The Human Revolution* (eds Mellars, P. & Stringer, C. B.) 62–108 (Princeton Univ. Press, Princeton, New Jersey, 1989).
- Thorne, A. G. & Wolpoff, M. H. *Scient. Am.* **266**, 76–83 (1992).
- Wolpoff, M. H. et al. *Science* **241**, 772–773 (1988).
- Smouse, P. E. & Long, J. C. *Yrbk phys. Anthropol.* **35**, 187–213 (1992).
- Sokal, R. R., Oden, N. L. & Wilson, C. *Nature* **351**, 143–145 (1991).
- Sneath, P. H. & Sokal, R. R. *Numerical Taxonomy* (Freeman, San Francisco, 1973).
- Mantel, N. *Cancer Res.* **27**, 209–220 (1967).
- Vandermeersch, B. *Les Hommes Fossiles de Qafzeh (Israel)* (CNRS, Paris, 1981).
- Stringer, C. B., Grün, R., Schwarcz, H. P. & Goldberg, P. *Nature* **338**, 756–758 (1989).
- Grün, R. & Stringer, C. B. *Archaeometry* **33**, 153–199 (1991).
- Valladas, H. et al. *Nature* **331**, 614–616 (1988).
- Schwarcz, H. P. et al. *J. hum. Evol.* **17**, 733–738 (1988).
- Mercier, N. et al. *J. arch. Sci.* **20**, 169–174 (1993).
- Waddle, D. M. thesis, State Univ. New York at Stony Brook (1993).
- Grün, R., Beaumont, P. B. & Stringer, C. B. *Nature* **344**, 537–539 (1990).
- Cooke, H. B. S., Malan, B. D. & Wells, L. H. *Man* **45**, 6–13 (1945).

- Hay, R. L. in *Results of the Laetoli Expeditions 1975–1981* (eds Leakey, M. D. & Harris, J. M.) 23–47 (Oxford Univ. Press, UK, 1987).
- Leakey, M. D. & Hay, R. L. in *L'Homo erectus et la Place de l'Homme de Tautavel Parmi les Hominides Fossiles* (ed. de Lumley, M. A.) 753–765 (1^{er} Cong. Int. Paleont. Hum., Nice, 1982).
- Clarke, R. J. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 301–305 (Liss, New York, 1985).
- Howells, W. W. *Papers of the Peabody Museum* **67**, 1–259 (1973).
- Vandermeersch, B. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 306–309 (Liss, New York, 1985).
- Stringer, C. B. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 289–295 (Liss, New York, 1985).

ACKNOWLEDGEMENTS. I thank W. Jungers, S. Leigh, L. Martin, R. Sokal, D. Stock and B. Thomson for comments and/or technical assistance, and the following for access to fossil specimens: G. Avery, B. Arensburg, L. Beck, J. S. Brink, R. J. Clarke, J.-J. Cleyet-Merle, S. Condemi, A. Czarnetzki, H. deLumley, M. Dockalova, D. Greene, H. E. Joachim, G. Koufos, R. Kritscher, A. Langaney, M. Leakey, R. Macchiarelli, G. Manzi, M. Mbago, J. Monge, A. Morris, H. Müller-Beck, R. Orban, D. Pilbeam, M. Soubeyran, C. B. Stringer, I. Tattersall, P. V. Tobias, J. Zias and R. Ziegler. This research was supported by grants from the LSB Leakey Foundation, Sigma Xi Foundation, the NSF and the Doctoral Program in Anthropological Sciences of SUNY, Stony Brook.

High resolution of human evolutionary trees with polymorphic microsatellites

A. M. Bowcock*, A. Ruiz-Linares†, J. Tomfohrde*, E. Minch†, J. R. Kidd‡ & L. L. Cavalli-Sforza†

* Department of Pediatrics, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75235-8591, USA

† Department of Genetics, Stanford University, Stanford, California 94305, USA

‡ Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

GENETIC variation at hypervariable loci is being used extensively for linkage analysis¹ and individual identification², and may be useful for inter-population studies³⁻⁵. Here we show that polymorphic microsatellites (primarily CA repeats) allow trees of human individuals to be constructed that reflect their geographic origin with remarkable accuracy. This is achieved by the analysis of a large number of loci for each individual, in spite of the small variations in allele frequencies existing between populations^{6,7}. Reliable evolutionary relationships could also be established in comparisons among human populations but not among great ape species, probably because of constraints on allele length variation. Among human populations, diversity of microsatellites is highest in Africa, which is in contrast to other nuclear markers and supports the hypothesis of an African origin for humans.

We examined 30 microsatellite loci in about ten individuals, from each of 14 indigenous populations chosen from the five continents. As a measure of similarity between the multiple locus genotype of two individuals, we calculated the proportion, P_s , of alleles that they share averaged over loci, where P_s is the number of shared alleles summed over loci/ ($2 \times$ number of loci compared). A distance measure between pairs of individuals was calculated as $(1 - P_s)$. A tree constructed from the pairwise inter-individual distances shows that individuals cluster according to their geographic origin (Fig. 1). Of the 148 individuals examined, 130 (87.8%) form discrete clusters that coincide with the continent of origin of the sample. The position of seven individuals (4.7%) is not clearly defined in the tree, and eleven individuals (7.4%) fall in a continental cluster that does not correspond to their geographic origin. Within continents these samples tend to form sub-clusters that correspond to their population of origin, with the exception of East Asian and some African populations. Nine out of 14 populations form sub-clusters that usually include more than 50% of the individuals from that population.

The tree shown in Fig. 1 contrasts sharply with trees derived from mitochondrial (mt)DNA in which it is rare for individuals of related geographic origin to form discrete clusters^{8,9}. The clustering pattern with microsatellites is not due to continent- or

population-specific alleles. Among the 342 alleles detected, there are 78 that are present in one continent only (43 in Africa, 14 in Oceania, 9 in East Asia, 6 in Europe, 6 in America), but they always have very low frequencies (4.1% on average). Furthermore, there is no indication of high continental differentiation: the average distance ($1 - P_s$) among individuals within the 14 populations is 0.64, versus an average of 0.73 among individuals from different continents, showing that the bulk of the genetic variation is within populations, as seen with other markers^{10,11}.

Construction of trees of individuals with samples of the loci indicates that there is no subset that is particularly informative. However, loci with higher F_{st} values (where for each locus, $F_{st} = \Sigma V_i / [\Sigma P_i(1 - P_i)]$; P_i , V_i are the mean and variance of gene frequencies of allele i summed over all L alleles at the locus) tend to produce trees where clusters are more defined. F_{st} is negatively correlated with locus heterozygosity ($r = -0.49$, $P < 0.01$), suggesting that loci with more extreme diversity and presumably higher mutation rates might on the whole be less informative.

Although most upper segments in the tree of Fig. 1 are very short, the various populations branch in an order similar to that seen in trees derived from classical markers or restriction-fragment length polymorphisms (RFLPs)^{12,13}. The first split separates Africans; this is followed by the close separation of Europeans, East Asians and Pacific populations. Finally, Americans branch off from Asian populations.

Population relationships were also examined using genetic distance measures based on allele frequencies. In the tree of populations (Fig. 2a) there is strong statistical support for the separation of African populations from other world populations (100% of bootstrap resamplings), in agreement with the hypothesis of an African origin for humans. The second split in this tree is supported by 84% of the bootstraps and separates Europeans from other populations, an observation that could reflect a real phylogenetic event, as suggested on the basis of non-DNA markers¹⁴, or could be due to Europeans having received genetic flow from both Africa and Asia¹⁵. Finally, with lesser statistical support (56% of bootstraps), Amerindians appear most closely related to East Asians, a classical observation but one that has been challenged recently¹⁴.

A subset of ten loci was also examined in the three primates closest to humans (chimpanzees, gorilla and orang). A genus tree constructed from the allele frequencies has very little structure, indicating that microsatellites do not provide adequate information for this type of comparison (Fig. 2b). Table 1 shows that the average length of alleles at the loci examined is quite similar for the four primates. Furthermore, although the non-human primate/human variance ratios are significantly higher than one for five loci, they are low considering that the divergence between primates took place at least 10 times (probably 100) earlier than the divergence among humans¹⁵. This suggests that there is a constraint on allele-length variation, as under an unconstrained

TABLE 1 Variation in allele length at ten microsatellite loci within humans and great apes

Locus	Mean allele length (base pairs)				Variance for non-humans	Variance for humans	<i>F</i> ratio* (d.f.) (non-human/human)
	Human	Chimp	Gorilla	Orang			
<i>ACTC</i>	83.1	78.0	75.9	75.7	21.8	470.88	0.31 (45,263)
<i>D13S133</i>	154.9	144.6	122.4	123.2	132.50	422.76	0.31 (45,265)
<i>D13S137</i>	111.5	105.6	113.5	99.4	93.61	48.53	1.93 (47,267)
<i>D13S227</i>	152.7	152.7	137.6	144.0	54.37	21.41	2.54 (47,259)
<i>D13S193</i>	136.6	134.3	140.6	116.0	123.40	47.37	2.61 (47,273)
<i>084XC5</i>	83.4	81.5	81.6	87.4	10.19	23.34	0.44 (45,265)
<i>D13S119</i>	132.5	129.0	135.5	124.7	42.46	34.69	1.22 (43,283)
<i>D13S118</i>	193.1	191.5	190.6	190.8	15.17	12.05	1.26 (39,281)
<i>D13S125</i>	147.4	141.4	148.3	141.0	120.57	49.08	2.46 (45,269)
<i>Utsw1523</i>	179.1	175.8	174.9	173.6	18.62	5.23	3.56 (47,259)

* Critical values for F at $P = 5\%$ are between 1.44 and 1.49.

TABLE 2 Genetic variation of human populations detected with different genetic markers (mean heterozygosity $\times 100$)

	110 Classical polymorphisms*	Nuclear RFLPs†	Microsatellites	Microsatellites‡	mtDNA§
Africa	16.3 $\pm$ 1.4	29.7 $\pm$ 0.7	80.7 $\pm$ 1.4	8.9 $\pm$ 0.4	2.32
Europe	20.2 $\pm$ 1.7	37.9 $\pm$ 1.5	73.0 $\pm$ 1.6	7.1 $\pm$ 0.5	0.95
Asia	18.9 $\pm$ 1.9	32.7 $\pm$ 1.2	68.5 $\pm$ 2.1	7.3 $\pm$ 0.4	1.65
Oceania	13.7 $\pm$ 1.8	27.5 $\pm$ 1.2	63.6 $\pm$ 1.8	7.4 $\pm$ 0.5	
America	15.5 $\pm$ 1.8	29.3 $\pm$ 3.8	58.8 $\pm$ 2.3	6.1 $\pm$ 0.4	1.29

An analysis of variance shows significant differences between the heterozygosities of the five continents detected with microsatellites ($F_{4,142} = 20$, $P < 0.01$). The difference between Africa and Europe is also significant ($t_{57} = 8$, $P < 0.01$).

* Data compiled elsewhere¹⁵.

† The values for Africa (two populations), Europe (one population), Asia (two populations) and Oceania (three populations) are calculated from data on 79 RFLPs (ref. 23 and Lin *et al.*, manuscript in preparation). Values for America (two populations from North America and three populations from South America) are based on 35 other RFLPs (unpublished data).

‡ Number of alleles detected per locus. Mean number of individuals typed per locus per continent: Africa (27.6), Europe (26), Asia (28.2), Oceania (27.4), America (27.4).

§ Nucleotide diversity (heterozygosity at the nucleotide level) $\times 100$ of the D-loop region²⁴.

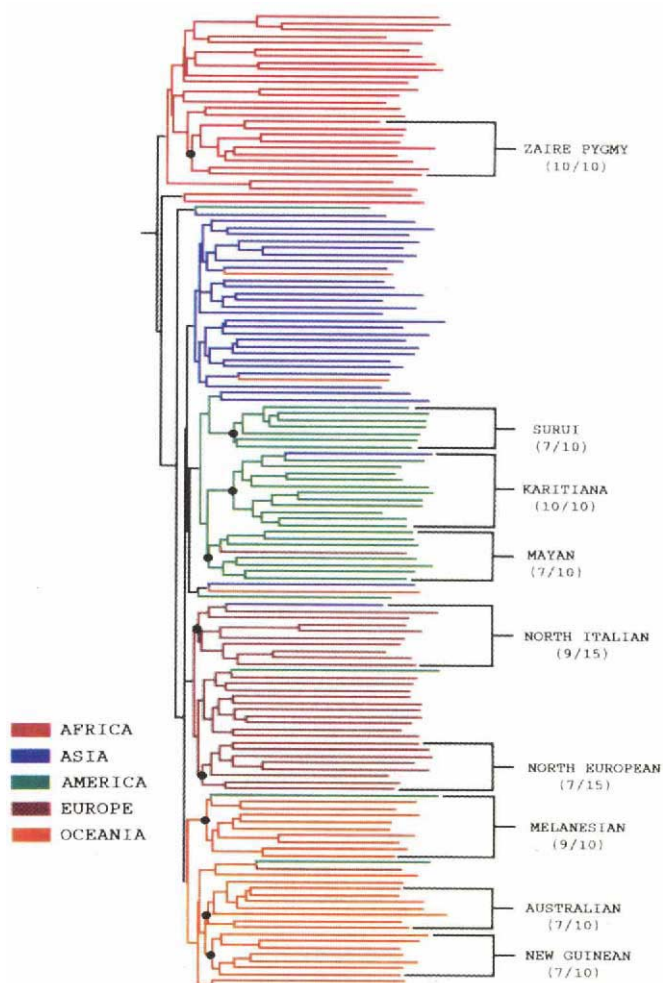
random walk the variance in allele size would increase linearly with time¹⁶. Although the details of the limitations to an unconstrained random walk for microsatellites, in which there is an absorbing or a reflecting barrier at the lower end^{17,18}, remain to be tested, it seems clear that there are limits to free variation.

In Table 2 the variation among human populations detected with microsatellites is compared with that detected with classical polymorphisms (blood groups and proteins), nuclear RFLPs and mtDNA. In the case of classical polymorphisms and nuclear RFLPs, Europeans show the highest level of variation, followed by East Asians, with Africans barely more heterogeneous than Australian aborigines or American Indians. By contrast, micro-

satellites show a significantly higher heterozygosity and number of alleles in Africa compared with other continents. Africa also has a higher diversity with mtDNA, but this finding is at present contested¹⁹. This observation supports the notion of an African origin for humans. A reasonable explanation for the discrepancy among different nuclear markers is the bias introduced by their initial selection in Europeans. This bias is likely to be less serious for markers with large numbers of alleles such as microsatellites.

The probability of distinguishing whether individuals belong to specific populations is probably enhanced in this material because the samples analysed here come from geographically discrete populations. This may generate apparent

FIG. 1 Neighbour-joining tree constructed from the pairwise distances between human individuals²⁵. A total of 148 samples from 14 human populations around the world were examined. The populations were as follows: three are from sub-Saharan Africa (Pygmies from Zaire and the Central African Republic, and the Lisongo population; two from Europe (North European and North Italian); three from East Asia (Chinese, Japanese and Cambodian); three from Oceania (Melanesian, New Guinean and Australian) and three from America (Maya from the Yucatan peninsula, Surui and Karitiana from the Amazon basin). A mean of 27.7 loci were typed on each individual (range 15–30), producing a unique multiple locus genotype for each individual. Distances between pairs of individuals were calculated from the proportion of alleles shared, as indicated in the text. There were no instances of pairs of individuals not sharing alleles. The loci typed by polymerase chain reaction (PCR) consisted of 29 CA repeats and one tetranucleotide repeat, all located on chromosomes 13 or 15. The exact location of two loci is not yet known, 8 other loci are located less than 1 cM away from their nearest neighbour and thus are likely to be in linkage disequilibrium, which with our small sample sizes is difficult to detect. The other 20 loci are located at an average distance of 8.8 cM from their nearest neighbour and presumably would have no detectable linkage disequilibrium except perhaps in very large samples. Methods for genotyping of microsatellites have been described²⁶. Coloured lines represent samples from different continents. Black dots indicate nodes that define population clusters. Numbers in parentheses are the fraction of individuals from the respective population found in a cluster. This tree and those shown in Fig. 2 were generated with programs in the PHYLIP package²⁷. The tree was rooted using midpoint rooting.



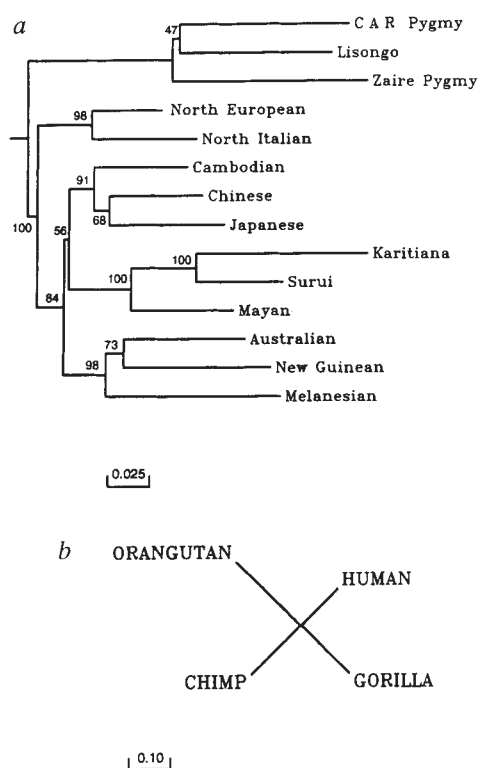


FIG. 2 *a*, Neighbour-joining tree relating the 14 populations examined. This tree was constructed using *f* (chord distance) as an estimate of genetic distance between populations²⁸. The numbers at the nodes are bootstrap values for 100 bootstrap resamplings²⁷ of the 30 loci typed. Other measures of genetic distance produced trees with identical topology but with slightly lower bootstrap values. The root of the tree was placed at the midpoint between the two most distantly related populations (Central African Republic (CAR) Pygmy and Karitiana). *b*, Neighbour-joining tree of the four primate species examined, based on chord distances between taxa (other distances produced the same results). This tree is based on a subset of ten loci typed in eight individuals (on average) for each of three other primate species (chimpanzees, gorillas and orangutans). All loci were polymorphic in all species, with the exception of locus D13S193, which was monomorphic in the orangutans tested. The scale is in distance units in both *a* and *b*.

discontinuity^{20,21} but, in general, human genetic geography shows high continuity^{15,22}. A geographically random sample would probably give a more complex and less discontinuous picture. Nevertheless, the results illustrate the resolving power of modern genetic analysis. □

Received 23 November 1993; accepted 24 January 1994.

- Weissenbach, J. et al. *Nature* **359**, 794–801 (1992).
- Pena, S. D. J., Chakraborty, R., Epplen, J. T. & Jeffreys, A. J. (eds) *DNA Fingerprinting: State of the Science* (Birkhauser, Basel, 1993).
- Gilbert, D. A., Lehman, N., O'Brien, S. J. & Wayne, R. K. *Nature* **344**, 764–767 (1990).
- Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T. & Chakraborty, R. *Genomics* **12**, 241–253 (1992).
- Chakraborty, R., Deka, R., Jin, L. & Ferrell, R. E. *Am. J. hum. Biol.* **4**, 387–397 (1992).
- Smouse, P. E., Spielman, R. S. & Park, M. H. *Am. Nat.* **119**, 445–463 (1982).
- Mitton, J. B. *Am. Nat.* **111**, 203–212 (1977).
- Cann, R. L., Stoneking, M. & Wilson, A. C. *Nature* **325**, 31–36 (1987).
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. *Science* **253**, 1503–1507 (1991).
- Lewontin, R. *Evol. Biol.* **6**, 381–398 (1972).
- Nei, M. & Roychoudhury, A. *Science* **177**, 434–436 (1972).
- Cavalli-Sforza, L. L., Piazza, A., Menozzi, P. & Mountain, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6002–6006 (1988).
- Bowcock, A. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 839–843 (1991).
- Nei, M. & Roychoudhury, A. *Molec. Biol. Evol.* **10**, 927–943 (1993).
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *A History and Geography of Human Genes* (Princeton Univ. Press, Princeton, in the press).
- Karlin, S. *A First Course in Stochastic Processes* (Academic, New York, 1969).
- Walsh, J. B. *Genetics* **115**, 553–567 (1987).
- Tachida, H. & Iizuka, M. *Genetics* **131**, 471–478 (1992).

- Templeton, A. R. *Am. Anthr.* **95**, 51–72 (1993).
- Barbujani, G. & Sokal, R. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1816–1819 (1990).
- Zei, G. et al. *Ann. hum. Genet.* **57**, 123–140 (1993).
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *Science* **259**, 639–646 (1993).
- Bowcock, A. M. et al. *Gene Geog.* **5**, 151–173 (1991).
- Horai, S. et al. *Molec. Biol. Evol.* **10**, 23–47 (1993).
- Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).
- Bowcock, A. et al. *Genomics* **15**, 376–386 (1993).
- Felsenstein, J. *PHYLIP* (Phylogeny Inference Package) Version 3.5c (Department of Genetics, Univ. Washington, Seattle, 1993).
- Cavalli-Sforza, L. L. & Edwards, A. W. F. *Evolution* **32**, 550–570 (1967).

ACKNOWLEDGEMENTS. This work was supported by grants from the NIH (to L.C.S. and K. K. Kidd), the Wenner Gren Foundation (L.C.S.), and the NSF (K. K. Kidd). B. Hewlett helped in obtaining the African cell lines. The North Italian cell lines were made by G. B. Ferrara in Genova with blood samples from Associazione Donatori di Sangue, Bergamo. DNA samples from New Guinea and Australia were provided by the late Allan C. Wilson. We thank the Lucille P. Markey Charitable Trust for funding (to L.C.S.) travel to Africa to collect samples, C. Dunn for technical help, and C. Campbell with help in preparation of the manuscript.

Bcl-2 expression promotes B- but not T-lymphoid development in *scid* mice

Andreas Strasser, Alan W. Harris, Lynn M. Corcoran & Suzanne Cory*

The Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

EXPRESSION of antigen receptors is vital for the development of B and T lymphocytes. In mice with the *scid* mutation^{1,2}, which are unable to make productive rearrangements of their immunoglobulin and T-cell receptor (TCR) genes, lymphopoiesis aborts at an early stage. The death of the immature lymphocytes by apoptosis³ is postulated to result from a failure to receive a survival signal induced by receptor engagement⁴. Consistent with this hypothesis, introduction of immunoglobulin or TCR transgenes into *scid* mice promoted an increase in B- or T-lymphoid cells, respectively^{5–7}. As the protein encoded by the *bcl-2* gene can inhibit cell death^{8,9}, we tested whether lymphopoiesis could be rescued in *scid* mice by crossing in a *bcl-2* transgene. Strikingly, the *bcl-2/scid* mice accumulated almost normal numbers of B-lymphoid cells which lacked surface immunoglobulin but expressed markers of maturity. T-cell development remained blocked. Introducing a TCR transgene enabled *bcl-2/scid* mice to develop normal numbers of CD4⁺8⁺ thymocytes even in the absence of immunological selection, suggesting that T cells become competent to respond to *bcl-2* protein only after the TCR complex is displayed at the cell surface.

The *bcl-2* protein (Bcl-2) is known to enhance the survival capacity of B and T cells at several stages of differentiation^{10–13}, but its effect in pro-B and pro-T cells has not previously been addressed. We introduced the Eμ-*bcl-2*-36 transgene, which is expressed in both lymphoid lineages¹², into *scid* mice by appropriate genetic crosses. The *scid* mice had the normal low number of pro-B and pro-T cells and essentially no mature lymphoid cells^{5,14}, whereas the *bcl-2/scid* mice had large numbers of cells expressing the B-lineage marker CD45R (B220) in the bone marrow, spleen and blood (Table 1), the total number approaching that found in normal mice. The increase was not due to augmented proliferation, because most of the cells were small and quiescent. Neither was it due to increased 'leakiness' of the *scid* mutation², because no immunoglobulin-bearing cells were detectable in the bone marrow (<10⁵ per femur), spleen (<10⁶ per organ), or peripheral blood (<3 × 10⁴ ml⁻¹) (panels I in Fig. 1a and b).

Unexpectedly, most B220⁺ cells in the *bcl-2/scid* mice had developed beyond the pro-B stage. Pro-B and early pre-B cells display various combinations¹⁵ of the markers Thy-1, S7(CD43),

* To whom correspondence should be addressed.

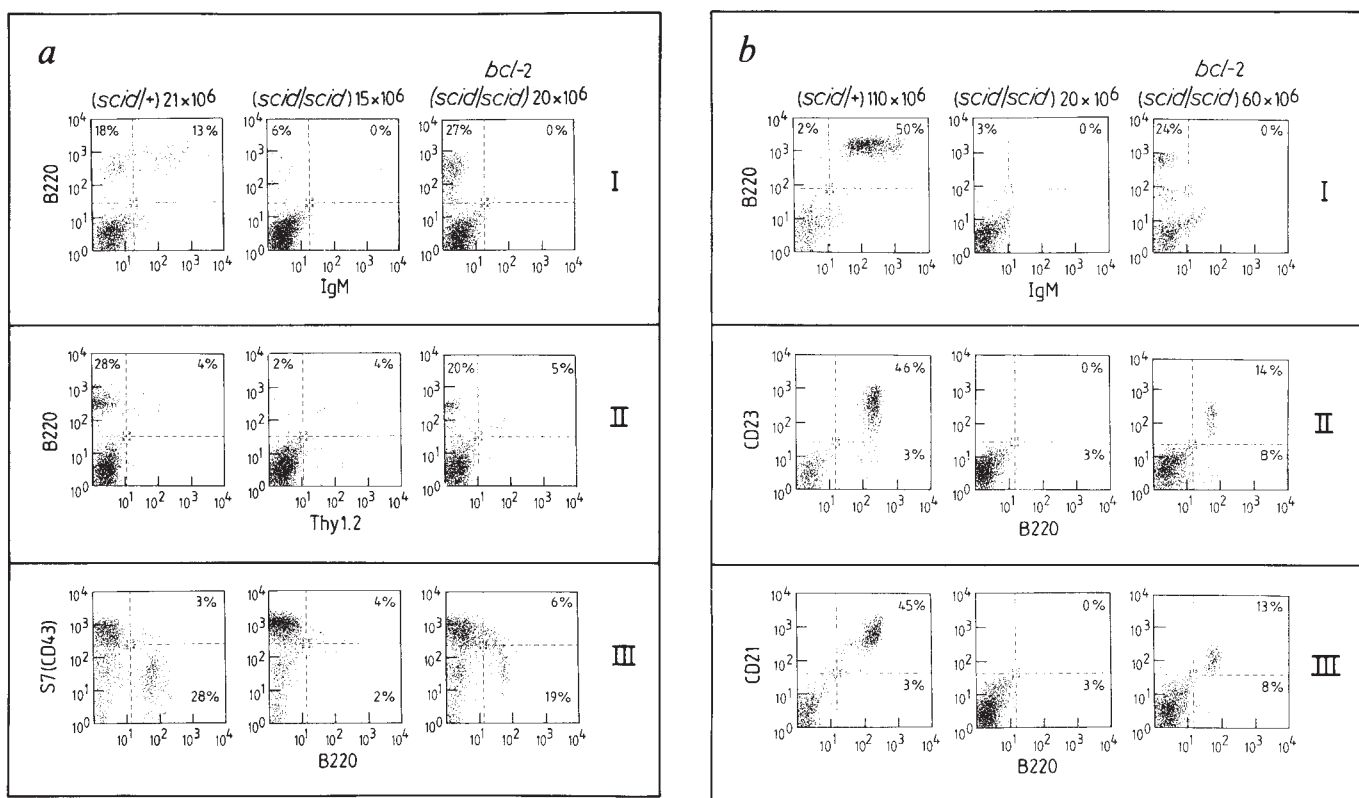


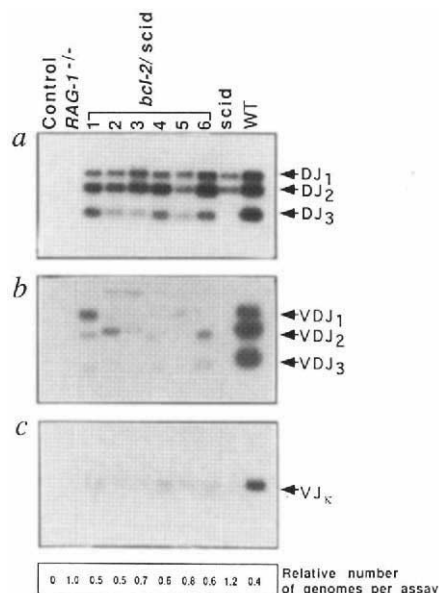
FIG. 1 Phenotypic analysis of B-lymphoid excess in *a*, the bone marrow, and *b*, spleen of *bcl-2/scid* mice. The B cells of non-*scid* *bcl-2* littermates expressed levels of the mature markers IgM, CD23 and CD21 (and IgD, CD22, CD40 and class II MHC molecules) like those in *scid/+* mice (data not shown).

METHODS. Nucleated cells from bone marrow and spleen were analysed by immunofluorescence and flow cytometry^{11,12} using a FACScan (Becton Dickinson). Representative dot-plots from one of at least 5 experiments are presented; 5,000–10,000 cells were analysed per sample. Staining with FITC-labelled antibodies is shown on the x-axis, biotinylated antibodies plus R-PE–streptavidin (Caltag) on the y-axis.

The monoclonal antibodies used were RA3-6B2 (anti-B220), B3B4 (anti-CD23), 5.1 (anti-IgM), 7G6 (anti-CD21), S7 (anti-CD43) and 30-H12 (anti-Thy 1.2). The $\epsilon\mu$ -*bcl-2*-36 transgenic mouse strain on a (C57BL/6 × SJL)_{F2} genetic background has been described¹². $\epsilon\mu$ -*bcl-2*-36 mice were serially crossed with homozygous C.B-17-*scid* mice¹ for two or more generations. The *bcl-2* transgene was detected by tail DNA dot-hybridization with an SV40 probe¹¹. As *scid* mice lack slgM^+ B cells², homozygosity for the *scid* mutation was diagnosed by FACS analysis of peripheral blood leukocytes after staining with FITC-labelled anti-IgM antibody and biotinylated anti-B220 antibody, followed by phycoerythrin–streptavidin.

FIG. 2 Increased frequency of immunoglobulin gene rearrangement in *scid* B-lymphoid cells expressing a *bcl-2* transgene. *a–c*. Assays for D μ to J μ (*a*) V μ to D μ (*b*) and V κ to J κ (*c*) rearrangements. The control is a polymerase chain reaction (PCR) buffer control lacking DNA; *rag-1*^{−/−} cells came from a homozygous mutant mouse generated by targeting the *rag-1* gene in embryonic stem cells (L.C., unpublished results). The unusual sizes of some of the VDJ amplification products presumably reflect the nature of the *scid* defect in gene rearrangement, in which deletions are generated during the joining process². WT, wild type.

METHODS. Sorted B220⁺ cells from bone marrow were lysed, extracts were prepared and PCR performed as described²⁸. The resulting gels were blotted and hybridized with germ-line J μ (*a* and *b*) or J κ (*c*) probes. To ensure that the number of templates in each assay was equivalent, a dilution series of each extract was prepared and PCR-assayed using primers for the *rag-2* gene (5' primer: GATGTCCTGAACCCAGATACGGC, 3' primer: CCTGGCAAGACTGTGCAATCACTGC). Dilutions in the linear range of the assay, giving similar signals on a stained gel, were used in the assay shown. The indicated relative number of genomes in each assay was calculated by counting radioactivity after hybridizing a *rag-2* probe to a Southern blot of the *rag-2* PCR assay using a PhosphorImager (Molecular Dynamics).



c-Kit, BP-1 and PB76 (refs 14, 16, 17). Although the *bcl-2/scid* mice contained a few such cells, most of the B220⁺ cells in the bone marrow and spleen lacked these markers (Fig. 1a, II and III, and data not shown). Instead, they displayed molecules normally characteristic of immunoglobulin-bearing B cells¹⁸⁻²⁰: CD23 (Fig. 1b, II) and the complement receptor CD21 (Fig. 1b, III) as well as CD22, CD40 and high levels of class II major histocompatibility complex (MHC) molecules (data not shown). The B-lymphoid cells from the *bcl-2/scid* mice had also undergone extensive immunoglobulin gene rearrangement. *V**D**J* heavy-chain and even κ light-chain gene recombination was frequent (Fig. 2b, c), whereas only *DJ* recombination occurred in the *scid* mice (Fig. 2a), as expected²¹. Both in phenotype and genotype, therefore, *bcl-2/scid* B-lymphoid cells are relatively mature, despite their lack of surface immunoglobulin.

The impressive ability of the *bcl-2* transgene to rescue B-lymphoid development in *scid* mice suggests that upregulation of Bcl-2 may be a signal for positive selection of immature B-lymphoid cells in the bone marrow. By analogy with findings for peripheral B cells in germinal centres²², it seems likely that Bcl-2 synthesis is induced as a result of engagement of the antigen receptor, thereby ensuring the survival of cells that have achieved productive rearrangement of receptor genes.

The *bcl-2* transgene had no discernible impact on T-cell development in *scid* mice. The thymus remained tiny (Table 1) and the thymocytes were all large, cycling pro-T cells which bore high levels of Thy-1 and either the α -chain of the interleukin-2 (IL-2) receptor or CD44 (data not shown), but no detectable CD3. Very few CD4- or CD8-bearing cells were evident in the thymus (<10⁵), spleen (<10⁶) or peripheral blood (<3 × 10⁴ ml⁻¹).

Consistent with the differential effect of the *bcl-2* transgene on B and T lymphopoiesis *in vivo*, B-lineage cells from *bcl-2/scid* mice survived much better in culture than those from *scid* littermates (Fig. 3a), and thymocytes of both genotypes died equally rapidly (Fig. 3b). The failure of the T-lymphoid cells to survive was not due to lack of expression of the transgene.

TABLE 1 Lymphoid cellularity of *bcl-2/scid* mice

Tissue and genotype	<i>bcl-2</i>	<i>n</i>	Cell content (×10 ⁶)		
			Total	B220 ⁺	Thy1 ⁺
Peripheral blood					
<i>scid</i> /+	—	4	4.6 ± 0.9	1.2 ± 0.3	1.9 ± 0.3
<i>scid/scid</i>	—	6	1.4 ± 1.0	0.02 ± 0.01	0.1 ± 0.1
<i>scid/scid</i>	+	5	2.6 ± 0.6	0.6 ± 0.2	0.2 ± 0.1
Bone marrow					
<i>scid</i> /+	—	5	17 ± 7	4.9 ± 2.4	
<i>scid/scid</i>	—	8	14 ± 3	1.0 ± 0.6	
<i>scid/scid</i>	+	8	13 ± 3	4.9 ± 1.6	
Spleen					
<i>scid</i> /+	—	5	120 ± 20	69 ± 12	25 ± 4
<i>scid/scid</i>	—	8	23 ± 9	2 ± 1	3 ± 2
<i>scid/scid</i>	+	8	100 ± 50	34 ± 14	7 ± 4
Thymus					
<i>scid</i> /+	—	5	250 ± 130		
<i>scid/scid</i>	—	8	3.5 ± 1.3		
<i>scid/scid</i>	+	8	3.0 ± 1.8		
TCR/ <i>scid/scid</i>	—	9	61 ± 26		
TCR/ <i>scid/scid</i>	+	7	210 ± 110		

Data shown are arithmetic means ± standard deviations from 3–9 individuals (*n*) of each genotype (produced as described in the legends to Figs 1 and 3) analysed at 5 to 10 weeks of age. Cell numbers are shown per ml for peripheral blood, per femur for bone marrow, and per whole organ for spleen and thymus. Immunofluorescence staining and FACS analysis are described in the legend to Fig. 1. Non-*scid* *bcl-2* littermate mice have a 2–4-fold excess (over normal mice) of B220⁺ cells and normal numbers of Thy-1⁺ cells¹² (data not shown).

Transgenic Bcl-2 protein was detected in sorted (Thy-1⁺) *bcl-2/scid* thymocytes by western blotting and flow cytometry at a level commensurate with that in more mature thymocytes from non-*scid* *bcl-2*-36 mice (data not shown), which do show enhanced survival (Fig. 3b, and ref. 12). Thus, inability to benefit

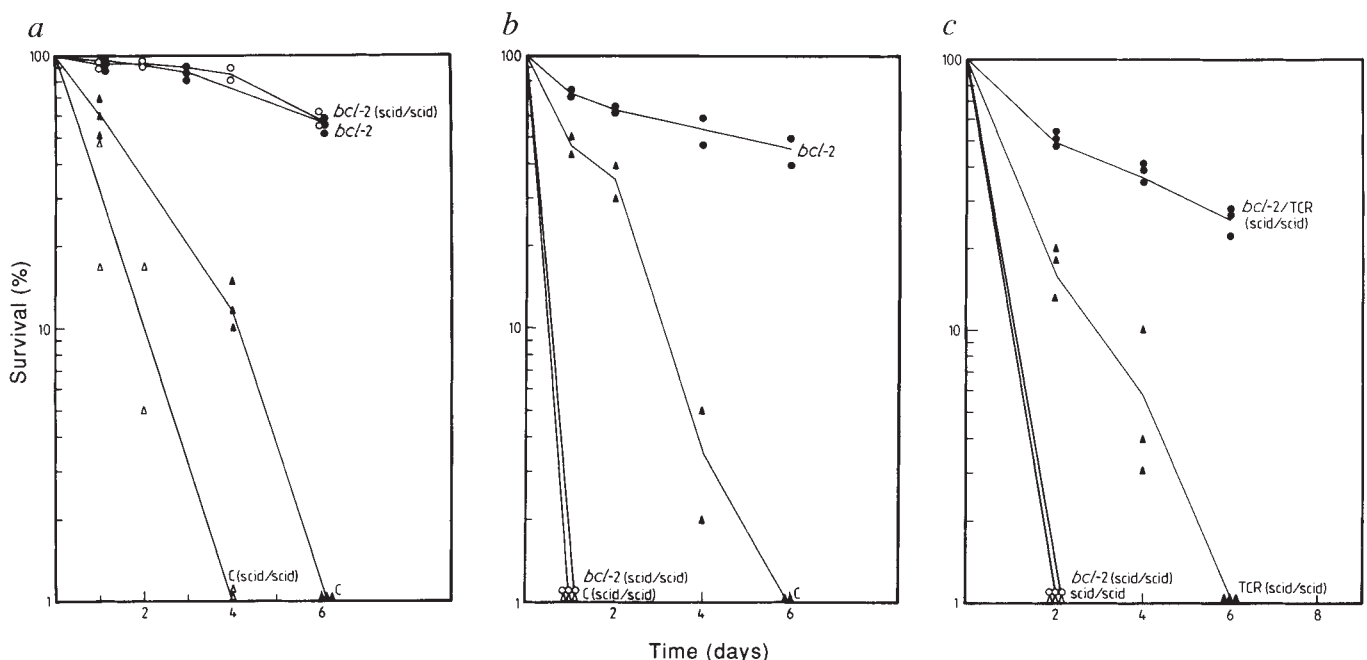


FIG. 3 Bcl-2 confers a survival advantage on *scid* lymphoid cells. *In vitro* survival of a, FACS-sorted B220⁺ bone marrow cells; b, thymocytes from *scid*/+ and *scid/scid* mice expressing the E μ -*bcl-2*-36 transgene and control (C) non-transgenic *scid*/+ and *scid/scid* littermates; c, thymocytes from female H-2D^{d/d} *scid/scid* mice expressing the anti-HY TCR transgene and/or the E μ -*bcl-2*-36 transgene.

METHODS. B220⁺slg⁻ bone marrow cells selected by flow cytometry

and thymocytes from 2–3 individual mice of each genotype were cultured and cell survival was measured by counting cells excluding trypan blue in a haemocytometer. E μ -*bcl-2*-36 mice homozygous or heterozygous for the *scid* mutation were produced as described in Fig. 1 legend. Similar mice bearing the anti-HY TCR were produced by crossing H-2^{d/d} *bcl-2/scid* mice with (H-2^{d/d} anti-HY TCR transgenic × C.B-17-*scid*)F₁ hybrids.

from Bcl-2 (at least at this level of expression) seems to be confined to very early T cells.

We have defined more precisely the stage of T-cell development at which Bcl-2 becomes effective by introducing the *bcl-2*-36 transgene into *scid* mice expressing an $\alpha\beta$ TCR transgene encoding reactivity against the HY antigen in association with H-2D^b (ref. 7). To avoid immunological selection by TCR engagement, we analysed anti-HY TCR/*scid* mice of H-2D^{d/d} haplotype, in which thymocytes are normally short-lived and not very numerous⁶. In this situation the *bcl-2* transgene increased the number of thymocytes by 3.5-fold²³ (Table 1) and greatly increased their ability to survive in culture (Fig. 3c). Almost all these cells (87%) expressed transgenic TCR α - and β -chains and were either CD4⁺8⁺ (76%) or CD4⁺8⁻ (20%), as previously reported for anti-HY TCR/*scid* mice⁷. Thus, immature T cells acquire responsiveness to Bcl-2 when they display the antigen-receptor complex, even in the absence of a specific TCR-MHC interaction.

Developing B and T cells might be expected to deploy the same basic strategy to acquire longevity. The ineffectiveness of the *bcl-2* transgene in *scid* pro-T cells therefore raises the possibility that a function necessary to supplement or complement the action of Bcl-2 is expressed later in the T lineage than the B lineage. A salient difference between pro-T and pro-B cells is that only the latter display the co-receptor which later associates with the mature antigen receptor. The B-lymphoid co-receptor, mb-1/B29 (Ig- α /Ig- β)²⁴, is present on pro-B cells in association with surrogate heavy (X) and light chains (VpreB and $\lambda 5$)²⁵. The T-lymphoid co-receptor CD3 is not expressed on pro-T cells; it only appears after rearrangement and expression of TCR genes permits the formation of CD3 complexed with β -gp33 (ref. 26) or $\alpha\beta$ heterodimers²⁷. Conceivably, therefore, a function essential for realization of cell survival-promoting activity of

Bcl-2 is induced by the co-receptor when it first appears at the cell surface. □

Received 13 October 1993; accepted 1 February 1994.

1. Bosma, C. G., Custer, R. P. & Bosma, M. J. *Nature* **301**, 527–530 (1983).
2. Bosma, M. J. & Carroll, A. M. *Rev. Immun.* **9**, 323–350 (1991).
3. Osmond, D. G. et al. *Blood* **79**, 1695–1703 (1992).
4. Rolink, A. & Melchers, F. *Cell* **66**, 1081–1094 (1991).
5. Reichman-Fried, M., Hardy, R. R. & Bosma, M. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2730–2734 (1990).
6. von Boehmer, H. A. *Rev. Immun.* **8**, 531–556 (1990).
7. Scott, B., Blüthmann, H., Teh, H. S. & von Boehmer, H. *Nature* **338**, 591–593 (1989).
8. Vaux, D. L., Cory, S. & Adams, J. M. *Nature* **335**, 440–442 (1988).
9. Nuñez, G. et al. *J. Immun.* **144**, 3602–3610 (1990).
10. McDonnell, T. J. et al. *Cell* **57**, 79–88 (1989).
11. Strasser, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8661–8665 (1991).
12. Strasser, A., Harris, A. W. & Cory, S. *Cell* **67**, 889–899 (1991).
13. Sentman, C. L., Shutter, J. R., Hockenbery, D., Kanagawa, O. & Korsmeyer, S. J. *Cell* **67**, 879–888 (1991).
14. Hardy, R. R., Kemp, J. D. & Hayakawa, K. *Curr. Top. Microbiol. Immun.* **152**, 19–25 (1989).
15. Li, Y.-S., Hayakawa, K. & Hardy, R. R. *J. exp. Med.* **178**, 951–960 (1993).
16. Müller-Sieburg, C. E., Whitlock, C. A. & Weissman, I. L. *Cell* **44**, 653–662 (1986).
17. Ogawa, M. et al. *J. exp. Med.* **174**, 63–71 (1991).
18. Waldschmidt, T. J., Conrad, D. H. & Lynch, R. G. *J. Immun.* **140**, 2148–2154 (1988).
19. Kinoshita, T. et al. *J. Immun.* **140**, 3066–3072 (1988).
20. Torres, R. M. et al. *J. Immun.* **149**, 2641–2649 (1992).
21. Pennycook, J. L., Chang, Y., Celler, J., Phillips, R. A. & Wu, G. E. *J. exp. Med.* **178**, 1007–1016 (1993).
22. Liu, Y. J. et al. *Eur. J. Immun.* **21**, 1905–1910 (1991).
23. Strasser, A., Harris, A. W., von Boehmer, H. & Cory, S. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1376–1380 (1994).
24. Reth, M. A. *Rev. Immun.* **10**, 97–121 (1992).
25. Melchers, F. et al. *Immun. Today* **14**, 60–68 (1993).
26. Groettrup, M. et al. *Cell* **75**, 283–294 (1993).
27. Ashwell, J. D. & Klausner, R. D. A. *Rev. Immun.* **8**, 531–556 (1990).
28. Schlissel, M. S., Corcoran, L. M. & Baltimore, D. *J. exp. Med.* **173**, 711–720 (1991).

Acknowledgements. We thank H. von Boehmer for anti-HY TCR mice; M. Stanley, J. Beaumont and F. Horsburgh for technical assistance; J. Parnis and K. Patane for animal husbandry; T. Kinoshita, E. Clark and S. Nishikawa for antibodies to CD21, CD22 and c-Kit; P. Lane for CD40 ligand; and J. M. Adams for discussion and critical review of this manuscript. A.S. is supported by fellowships from the Leukemia Society of America and the Swiss NSF. L.M.C. is a Cancer Research Institute Investigator, and S.C. is an International Research Scholar of the Howard Hughes Medical Institute. This work was supported by the National Health and MRC of Australia and the US National Cancer Institute.

Early determination of a mouse somatosensory cortex marker

Michel Cohen-Tannoudji*, Charles Babinet* & Marion Wasseff†

U106 INSERM, Hôpital de la Salpêtrière, 47 Bd de l'Hôpital, 75651 Paris Cedex 13, France

* Génétique des Mammifères, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

THE mammalian neocortex is subdivided into functionally distinct areas differing in cytoarchitecture and connectivity. Areal specification is thought to occur late in development^{1–3} and to be controlled by extrinsic cues, particularly thalamic afferents⁴. We have produced a transgenic mouse line in which β -galactosidase expression in the neocortex is largely restricted to layer-IV neurons of the somatosensory area. Transgene expression in these mice may be considered as an intrinsic marker of a somatosensory cortex identity. We investigated whether the fate of pieces of embryonic cortex from transgenic embryos is modified after transplantation to ectopic locations. Parietal or occipital cortex obtained on embryonic days 14–16 maintained their characteristics with respect to transgene expression after heterotopic transplantation to the cerebellum or neocortex of newborn hosts. This shows that the specification of neocortical areas involves a previously unsuspected early step of areal determination.

The H-2Z transgenic lines were generated using regulatory elements from a major histocompatibility complex class I gene

linked to the *Escherichia coli lacZ* gene⁵. Unexpectedly, the pattern of β -galactosidase expression, detected histochemically, was controlled by regulatory elements that were specific for each insertion site and varied between lines. The H-2Z1 line was analysed further because of the striking pattern of transgene expression outlining the somatosensory area of the neocortex (Fig. 1a). Staining with X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactosidase) was first detected in the parietal cortex by post-natal day 2 (P2) in two thin convergent stripes of cells (not shown). From P6 onwards, the X-gal-stained cortical area became wider and developed a patchy organization. The somatosensory area contains an orderly representation of the body surface^{6,7} and this pattern was precisely outlined by the blue X-gal reaction product in the cortex of H-2Z1 mice (Fig. 1a). X-gal staining was most intense between P6 and P14. In particular, the barrelfield is known to contain discrete representations of the sensory vibrissae⁸. These appeared as distinct blue patches in the transgenic animal (Fig. 1a). The X-gal-labelled cells were located in layer IV (Fig. 1b) and contained one or two blue spots near the cell nucleus. In addition, a small and individually variable number of blue cells were found in the deep layers of all cortical regions (Fig. 1b). Cells expressing β -galactosidase were also found in the pons, thalamus, hippocampus fascia dentata and cerebellum⁵.

To determine whether transgene expression can be considered as an intrinsic property of the somatosensory cortex, pieces of embryonic cortex were dissected out during the period of generation of layer-IV neurons (embryonic days 14–16 (E14–E16); ref. 9 and Fig. 1c) and transplanted in various ectopic locations. Two to three weeks after transplantation, the grafted brains were analysed for the presence of β -galactosidase-expressing cells by *in toto* X-gal staining (Table 1). When transplanted into the cerebellum of P6 hosts, pieces of E14 parietal and occipital cortex maintained their presumptive expression of the transgene. Grafts of parietal cortex contained a high density of blue cells

† To whom correspondence should be addressed at CNRS. URA 1414, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France.

TABLE 1 Pattern of transgene expression in ectopic cortical grafts

Origin of transplant	Host age	Site of transplantation	X-gal staining		
			Dense (%)	Scattered (%)	None (%)
E14 parietal	P6	Cerebellum	73 (8)	18 (2)	9 (1)
E14 occipital	P6	Cerebellum	0	0	100 (9)*
E14-E16 parietal	P0-P1	Occipital cortex	76 (44)	4 (8)	10 (6)
E14-E16 occipital	P0-P1	Parietal cortex	0	19 (7)	81 (30)

Transgene expression was analysed on *in toto* stained brains (see Fig. 2 methods). The success of transplantation was checked on sections of the grafted brains processed for ^3H -thymidine autoradiography. Fifty (83%: E15 (36) or E16 (14)) transplants were recovered from 60 brains that received E15-E16 grafts in the cortex. In the cerebellum, 95% of the E14 grafts were recovered. Numbers in parentheses are the number of grafted animals.

* All animals except one contained a transplant.

(Fig. 2a and b), whereas grafts of occipital cortex contained no transgene-expressing cells and could only be localized on counterstained sections (Fig. 2c). Thus, in a non-cortical environment, pieces of E14 cortex develop an expression of the transgene that is appropriate for their site of origin.

Several studies (reviewed in ref. 1) have indicated that many areal features of the cerebral cortex, including connectivity and cytoarchitecture, are governed by local cues. We therefore examined whether transgene expression could be modified in a new cortical environment. Pieces of E14-E16 parietal (Fig. 2d and e) and occipital (Fig. 2f-h) H-2Z1 cortex were grafted, respectively, into the occipital and parietal cortex of P0-P1 mice and analysed as before. More than 75% of the grafts maintained their presumptive expression of the transgene in a foreign cortical environment (Table 1). To distinguish host from grafted tissue, E15-E16 donors were labelled with ^3H -thymidine before transplantation. When their brains were treated with X-gal, sectioned and processed for autoradiography, X-gal staining was indeed localized to the ^3H -thymidine-labelled grafts. In addition, the X-gal-negative grafts could be unambiguously delineated and contained very few blue cells (Fig. 2h).

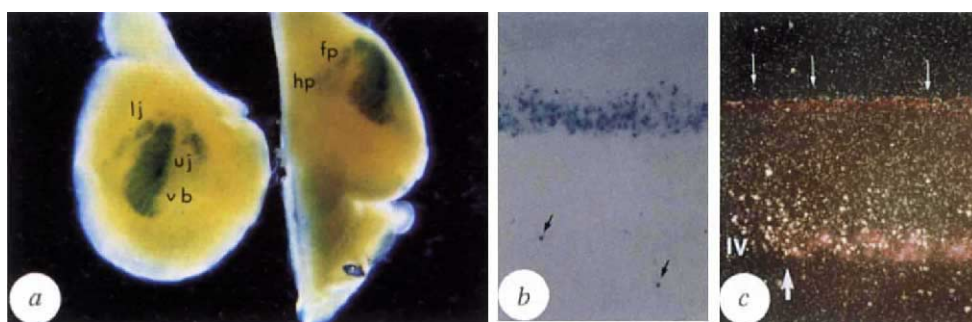
Because transplantation of H-2Z1 cortex into the cerebral cortex and cerebellum gave similar results, we investigated whether the transplants became integrated into their new cortical environment. When transplanted into the occipital cortex of a newborn host, the E17 rat parietal cortex can establish and maintain a projection to the host superior colliculus¹⁰. We repeated these experiments, using E14 H-2Z1 mice as donors and Fluorogold as a fluorescent retrograde axonal tracer. In addition to transgene-expressing cells, the grafts contained neurons, pre-

sumably layer-V neurons, which were filled with bright dots of Fluorogold (Fig. 2i, j). This indicated that their axons had reached the injection site in the host superior colliculus. Thus, even when connected with distant parts of the host brain, parietal grafts do maintain a 'somatosensory' fate, as indicated by transgene expression. Thus, in the H-2Z1 line, as early as E14, the parietal cortex is fated to express a specific marker of the somatosensory area.

It was possible that the somatosensory area identity was controlled by early-arriving thalamic axons^{11,12}. Thus, the thalamo-cortical projection was traced by inserting crystals of the fluorescent dye DiI into the dorsal thalamic region of aldehyde-fixed E14-E15 H-2Z1 brains. At E14, the DiI-labelled fibres were traced to the middle of the striatum (Fig. 3a), where they ended tipped by large growth cones (Fig. 3b). At E15, the thalamo-cortical axons reached the cortex, where they remained confined to a narrow region near the ventricular angle (Fig. 3c). Thus, the determination of transgene expression does not involve an interaction with thalamic axons before graft removal.

The notion that, in the rodent neocortex, area-specific features are not specified until late in development is supported by transplantation studies^{1,4,10} showing that pieces of developing rat neocortex can adapt their connections and some cytoarchitectural features to a new local environment, even at late fetal stages. However, such experiments do not exclude the existence of early intrinsic differences between neocortical areas. Although the differences in transgene expression that distinguish cortical areas in H-2Z1 mice are seen postnatally, our results strongly indicate that the mouse neocortex is already regionalized at E14, before the ingrowth of thalamic afferents, when layer-IV neuron pro-

FIG. 1 β -Galactosidase activity delineates the somatosensory area in one-week-old H-2Z1 mice brains stained *in toto* (a, c) or on frozen sections (b). The somatosensory cortex contains an ordered representation of body parts. a, Lateral (left) and dorsal (right) aspects of the right cerebral hemisphere of H-2Z1 mice. The pattern of staining corresponds to the representations of the vibrissae (vb), the upper and lower jaws (uj, lj) and the fore- and hind-paws (fp, hp). b, Most X-gal-labelled neurons are in layer IV; in addition, a scarce population of blue cells

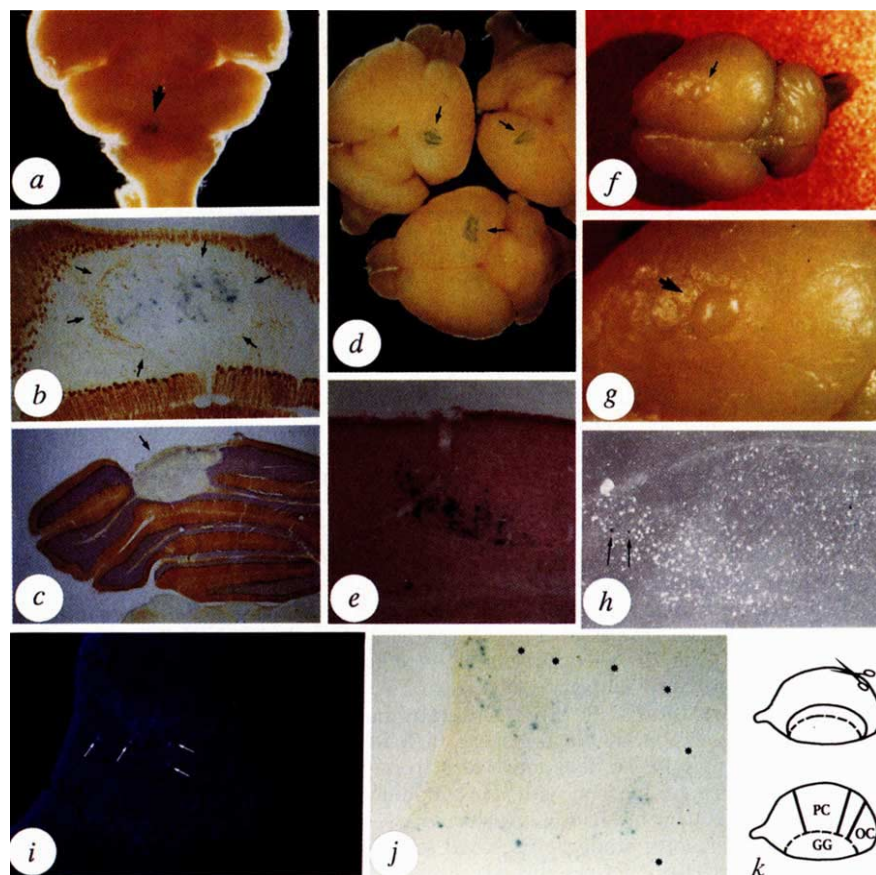


(arrows) is observed in deeper layers throughout the cortex. c, Coronal frozen section involving the somatosensory area (on the right side of the large arrow) of a P8 brain. The somatosensory cortex layer-IV neurons incorporated ^3H -thymidine injected at E15. The section was photographed under dark-field microscopy. X-gal-stained layer-IV neurons of the somatosensory cortex (IV) appear pink and the nuclei that have incorporated ^3H -thymidine are covered with bright dots. The thin arrows mark the pial surface.

METHODS. The H-2Z1 transgenic line was produced by microinjection into (C57 $\times$ SJL) F_2 fertilized eggs of a construct containing 2 kilobases of the H-2K^b promoter fused to the *E. coli* β -galactosidase reporter

gene⁵. Embryos (E15-E19) and pups (P0-P21), recovered from (C57 $\times$ CBA) F_1 females mated to transgenic males, were fixed by transcardiac aldehyde perfusion. The brains were dissected out and assayed *in toto* or on frozen sections for the histochemical detection of β -galactosidase activity as described⁵. The day of vaginal plug was considered to be E1, and the day of birth P0. Three pregnant females received an intraperitoneal injection of ^3H -thymidine ($5 \mu\text{Ci g}^{-1}$, specific activity 5 Ci mmol^{-1}) on day 15 of gestation. Their offspring were fixed at P8. The brains were incubated *in toto* with X-gal, then frozen-sectioned, mounted on gelatin-subbed slides and processed for autoradiography.

FIG. 2 Pieces of E14–E15 H-2Z1 embryonic cortex maintained their presumptive expression of the transgene when grafted in various ectopic locations: cerebellum (a–c) or cerebral cortex (d–h). E14 parietal (a, b) and occipital (c) H-2Z1 cortex transplanted into the cerebellum of P6 hosts were incubated with X-gal and examined on whole brains (a) or on sections (b, c) 2 weeks after grafting. An antiserum against calbindin was used to stain Purkinje cells, which appear brown in the host cerebellum (b, c). Grafts of parietal cortex contained β -galactosidase activity (blue staining in a and b), whereas no transgene-expressing cells were observed in grafts of occipital cortex. The presence of the occipital cortex transplants was checked on cresyl violet-counterstained sections of the grafted cerebella (c). d, e, β -Galactosidase expression in grafts of E14 H-2Z1 parietal cortex transplanted to the occipital cortex of newborn hosts. The stripes of X-gal-labelled cells are discernible in the grafts on whole brains reacted *in toto* (arrows in d) as well as on sagittal frozen sections counterstained with pararosaniline (e). f–h, In most cases, E15 occipital cortex transplants placed in the parietal cortex of newborn hosts (f, g; arrows) are devoid of X-gal labelling. h, The occipital transplants are identified by the presence of a dense population of cells that have incorporated 3 H-thymidine. Occasionally a few blue cells are observed in these grafts (arrows in h). i and j, Section through a parietal cortex transplant viewed under ultraviolet fluorescence optics (i) and bright-field microscopy (j). The pial surface is on the left. The host superior colliculus had been injected with Fluorogold 4 days before fixation and the brain was assayed for β -galactosidase activity before sectioning. The U-shaped transplant (outlined by stars in j) contained a row of transgene-expressing neurons which exhibit a faint green fluorescence under dark-field microscopy (i) and are stained in blue under bright-field (j). Presumptive layer-V neurons (i, arrows) contained bright dots of Fluorogold. k, See below.



METHODS. E14–E16 embryos were obtained from (C57 × CBA) F_1 females mated with heterozygous H-2Z1 males. E15–E16 embryos were labelled with 3 H-thymidine ($5 \mu\text{Ci g}^{-1}$, specific activity 5 Ci mmol^{-1}) 1 or 2 days before embryo recovery. The brains were dissected in sterile PBS containing 33 mM glucose. Transgenic embryos were recognized and selected for transplantation by the presence of a cluster of β -galactosidase-expressing cells in the pontine region. A tissue slice containing the pontine region was freed from meninges, incubated with a fluorescent β -galactosidase substrate (FDG; Molecular Probes) 18 and observed under FITC optics. The presence of the transgene was confirmed by polymerase chain reaction in the case of E14 embryos. The cerebral hemispheres were freed from meninges. As illustrated in k, the primordia of the medial cortex and hippocampus were discarded; the presumptive parietal (PC) and occipital cortices (OC) were delimited

by transversal cuts, then separated along the limit (dashed line) of the ganglionic eminence (GG). The skin over the head of cold anaesthetized non-transgenic newborn hosts was deflected and a hole was made in the skull using a syringe needle. A piece ($0.2\text{--}0.5 \text{ mm}^2$) of transgenic cortex was aspirated through a glass micropipette ($200 \mu\text{m}$ internal diameter). The pipette was inserted beneath the pia and moved, under visual control, parallel to the surface of the brain. The graft was ejected by blowing, while removing the pipette. Transplants were injected into the P6 cerebellum through a small incision over the cisterna magna. The skin was glued with cyanoacrylate and the pups were returned to their mothers. They were killed 17–28 days later, fixed by perfusion and their brains were assayed *in toto* for the histochemical detection of β -galactosidase. The brains were inspected for the presence of a graft under a stereomicroscope. They were infiltrated in sucrose, frozen-cut, processed for autoradiography, and counterstained with pararosaniline. Some animals bearing E14 parietal cortex transplants were anaesthetized 3 weeks after grafting and received an iontophoretic injection of Fluorogold (2% in saline) into the superior colliculus. After 3–4 days' survival they were fixed and processed as above.

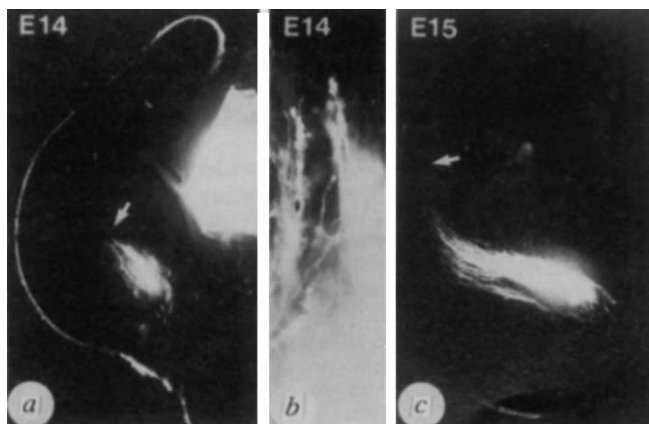


FIG. 3 Dil tracing of the developing thalamocortical projection in E14–E15 H-2Z1 embryos. At E14 (a, b), the fluorescent thalamocortical axons labelled by insertion of Dil into the dorsal thalamus ended midway into the striatal region (arrow in a) tipped with large growth cones (b). The thalamocortical axons reached the vicinity of the somatosensory cortex at E15 (arrow in c).

METHODS. The brains of E14–E15 H-2Z1 embryos were dissected in PBS and fixed by immersion in phosphate buffer (0.12 M, pH 7.4) containing 4% paraformaldehyde, 0.2% glutaraldehyde overnight at 4°C . A concentrated solution of Dil (D282; Molecular Probes) in ethanol or dimethyl formamide was dried on the surface of glass needles that were inserted into the dorsal thalamus of fixed embryos. The brains were left in the fixative at 30°C for 3 weeks. Coronal 100- μm thick Vibratome sections were mounted in glycerol and photographed under rhodamine fluorescence optics.

duction is still in progress. The early commitment of cerebral cortex neurons to the expression of position-specific markers has already been reported^{13–15}. The lateral and dorsomedial portions of rat neocortex can be distinguished as early as E13 by differences in the production of PC3.1-immunoreactive neurons¹³, and neurons from the rat limbic cortex are committed to express LAMP^{14,15} by E14. Our data extend these observations by showing the first evidence of an early regional commitment to the expression of an areal marker. Our grafting experiment does not examine how transgene-expressing cells later become strictly localized to the somatosensory area. Thalamic afferents are probably involved in the transition from regional to areal specification^{12,16,17}. Their role in this process is under investigation. □

Received 28 September 1993; accepted 21 January 1994.

1. O'Leary, D. D. M., Schlagger, B. L. & Stanfield, B. B. *Expl. Neurol.* **115**, 121–126 (1992).
2. Shatz, C. J. *Science* **258**, 237–238 (1992).
3. Walsh, C. & Cepko, C. L. *Science* **255**, 434–440 (1992).
4. Schlagger, B. L. & O'Leary, D. D. M. *Science* **252**, 1556–1560 (1991).
5. Cohen-Tannoudji, M., Morello, D. & Babinet, C. *Molec. Reprod. Dev.* **33**, 149–159 (1992).
6. Woolsey, C. N. in *Biological and Biochemical Bases of Behavior* (eds Harlow, H. F. & Woolsey, C. N.) 62–82 (Univ. Wisconsin Press, Madison, 1958).
7. Dawson, D. R. & Killakey, H. P. *J. comp. Neurol.* **256**, 246–256 (1987).
8. Woolsey, T. A. & Van der Loos, H. *Brain Res.* **17**, 205–242 (1970).
9. Fairén, A., Cobas, A. & Fonseca, M. J. *comp. Neurol.* **251**, 67–83 (1986).
10. O'Leary, D. D. M. & Stanfield, B. B. *J. Neurosci.* **9**, 2230–2246 (1989).
11. Catalano, S. M., Robertson, R. T. & Killakey, H. P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2999–3003 (1991).
12. Kennedy, H. & Dehay, C. *Cerebral Cortex* **3**, 171–186 (1993).
13. Arimatsu, Y. et al., *Proc. natn. Acad. Sci. U.S.A.* **89**, 8879–8883 (1992).
14. Barbe, M. F. & Levitt, P. J. *Neurosci.* **11**, 519–533 (1991).
15. Ferri, R. T. & Levitt, P. *Cerebral Cortex* **3**, 187–198 (1993).
16. Rakic, P. *Science* **241**, 170–176 (1988).
17. Rakic, P., Suner, I. & Williams, R. W. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2083–2087 (1991).
18. Nolan, G. P., Fiering, S., Nicolas, J.-F. & Herzenberg, L. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2603–2607 (1988).

ACKNOWLEDGEMENTS. We thank D. Jassik-Gerschenfeld and H. Kennedy for help and encouragement, D. Jassik-Gerschenfeld, H. Kennedy, J. Price, A. Prochiantz, P. Rakic, S. Sockanathan, P. Baldacci and C. Sotelo for critical reading of the manuscript, and P. Marchand for maintaining the transgenic lines.

Dynamics of brain activation during picture naming

R. Salmelin, R. Hari, O. V. Lounasmaa & M. Sams

Low Temperature Laboratory, Helsinki University of Technology,
02150 Espoo, Finland

THE cerebral representation of language, deduced from observing patients with brain lesions and from stimulations and recordings performed during brain surgery^{1,2}, has been further clarified by recent positron emission tomography³ and functional magnetic resonance imaging measurements⁴. We now expand this static view into the dynamics of cortical activation using the accurate spatio-temporal resolution of whole-head magnetoencephalography⁵. During picture naming, the conversion from visual to symbolic representation progressed bilaterally from the occipital visual cortex towards temporal and frontal lobes. Overt naming elicited the most widespread cortical activation. Some language-related sites also reacted, though more weakly or after a longer delay, during covert naming and even passive viewing.

Black-and-white line drawings of everyday objects (for example, vase, book, cat) were presented to six healthy right-handed subjects (age 25–34 yr; 4 males) for 100 ms once every 5 s, randomly to the lower quadrant of the left or right visual field. The subject first fixated on a point in the middle of the screen and ignored the stimuli, unaware of the tasks to follow ('passive viewing'). In the subsequent sessions he/she named the objects aloud first ('overt naming') and then silently ('covert naming'). In each condition, 80–100 stimuli were presented.

Magnetic brain signals, recorded with a whole-head

neuromagnetometer⁶, were averaged, time-locked to the appearance of the visual stimulus or to the onset of speech. Activated areas in the brain were sought for 1.5 s after stimulus onset. Evoked responses to visual, auditory and somatosensory stimulation provided functional landmarks to indicate the sites of sensory projection cortices⁷ (Fig. 1).

Figure 1 shows the spatiotemporal activation sequence for subjects S4 and S6 during overt naming of pictures presented to the left visual field. The right occipital visual area (yellow) reacted first in both subjects and was followed by early bilateral signals close to the temporo-parieto-occipital junctions (orange). In S4, left frontal activity (orange) occurred simultaneously; bilateral sites in the inferior part of the frontal lobes (red) then became active, followed by a site close to the vertex (red). In S6, an early signal lateral to the right somatomotor hand area (yellow) was followed by that from the left fronto-temporal cortex (orange). A late parietal source above the right auditory cortex (red) then became active.

Figure 2 illustrates data from all subjects, displayed on a schematic two-dimensional head. Based on the site and latency of activation, clusters of sources (inside dashed borders) were tentatively marked A–F. The earliest activation occurred in the posterior occipital cortex, contralateral to the stimulated visual field (area A; yellow). Within the next 200 ms, cortex close to the temporo-parieto-occipital junction in both hemispheres became involved (B1 and B2; orange), suggesting a contribution from the Wernicke's area/angular gyrus region and from the right-hemisphere counterpart. At about the same time, distinct signals emerged close to the left auditory cortex (C1; orange); responses from the homologous area in the right hemisphere (C2; red) peaked 100–200 ms later, independent of the side of stimulation. About 0.5 s after stimulus onset, activation increased close to the face motor representation area and in the region anterior to it (D1; red), probably representing Broca's area; the homologous cortical sites in the right hemisphere were activated within the same time interval (D2; red).

During overt vocalization, activation of D1 was immediately followed by signals produced close to the vertex (Fig. 2, F; red), arising from the vicinity of the supplementary motor area (SMA). Bilateral posterior frontal sites, midway to vertex (E1 and E2; orange), reacted within the first 400 ms, yet more consistently in the left than in the right hemisphere, presumably associated with motor preparation for mouth movements.

The left fronto-temporal region generated signals 400–600 ms after stimulus onset in every subject. Other source clusters consisted of data typically from three subjects. The combination of activated cortical sites was different for all subjects. The sources were located in both hemispheres, but the left-side sites became active earlier and were more strongly and consistently sensitive to naming. The wide variation in source locations between individuals agrees with intra-operative cortical recordings, in which language-specific sites were widely distributed in the left frontal, temporal and parietal lobes².

Figure 3 depicts the temporal behaviour of the active areas during the three experimental conditions in subjects S3 and S4. Early occipital visual responses (site A) did not systematically depend on the task, and signals evoked by passive viewing in sites B1 and B2 increased only slightly with naming. Signals not intensified during naming are apparently purely visual responses. Naming-specific signals were observed in areas C1, D1 and D2.

Averaging the data with respect to the onset of speech, 750–975 ms after picture onset, accentuated signals over areas E1, E2 and F, thus relating them to the production of speech, rather than to the processing of visual input. Picture presentation to the left or the right visual field did not produce any systematic differences in the responses of areas B–F.

Our results show that signal processing during picture naming advances from the posterior visual cortices to language- and eventually vocalization-related areas. This conclusion is supported by the spatiotemporal reactivity of cortical spontaneous

FIG. 1 Left and right views on the magnetic resonance image surface renditions of subjects S4 and S6; the sensory projection cortices (A, auditory; V, visual; S, hand somatosensory), derived from evoked response data, are indicated with black squares. The spherical symbols, colour-coded according to latency of activation, represent the three-dimensional locations of activated cortical sites during overt naming, projected onto the surface of the brain. The size of the symbol exceeds the 95% confidence radius (± 4 mm on the average) at all sites, except in the left frontal cortex of S4, where the confidence limits are indicated (± 9 mm; cross).

METHODS. Magnetic fields over the whole scalp were recorded with a helmet-shaped Neuromag-122 SQUID-magnetometer⁶. The data were analog low-pass filtered at 0.1 kHz and digitized at 0.3 kHz. The planar detectors produce maximum signal just above the active brain area. From visual inspection of the magnetic field distribution, a subset of channels was selected around the maximum. The three-dimensional location, strength and direction of the equivalent current dipole, characterizing the local activated brain area, was then obtained from a least-squares fit. Only sources accounting for $>85\%$ of the field variance were accepted. Several temporally or spatially distinct current dipoles were similarly identified. Subsequently, the single sources were introduced simultaneously into a multi-dipole model. All 122 channels were then included and the strengths of the dipoles were allowed to vary as a function of time to explain the measured field pattern and to verify the alternating dominance of individual source areas. (For a full description of the magnetoencephalography method, see ref. 5.)

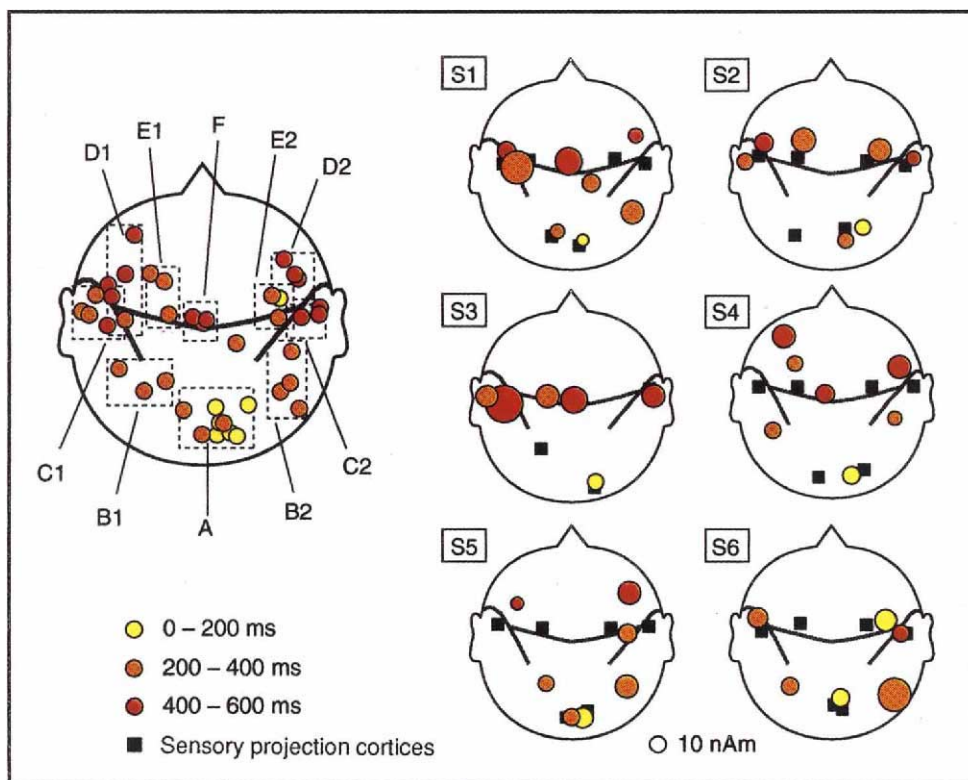
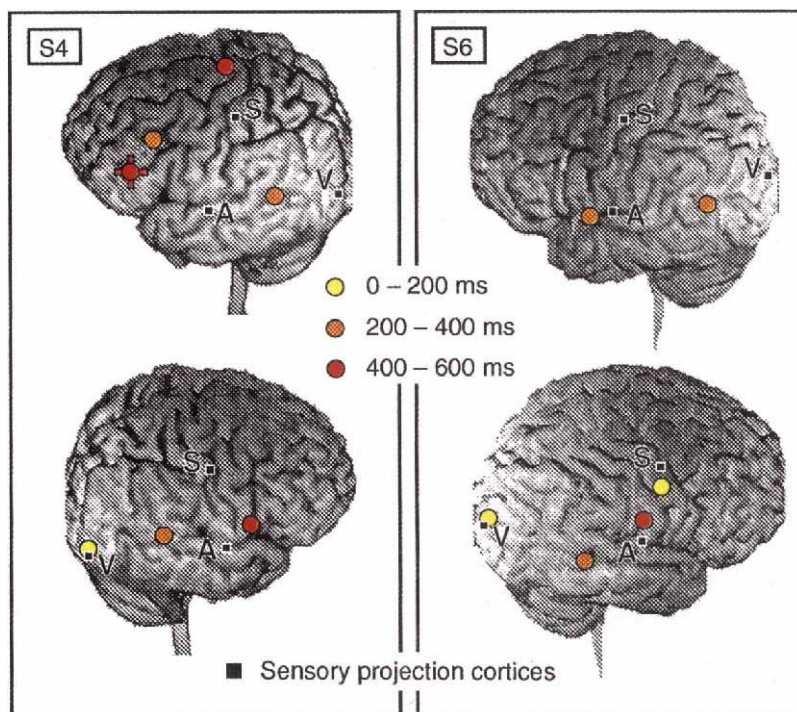


FIG. 2 Left, spatial and temporal relations of sources of evoked responses during an overt naming task for all six subjects, superimposed on a single schematic head. The activated sites, localized in three dimensions, were projected onto the head surface, which was subsequently flattened into a two-dimensional view for easy visualization. The heavy lines indicate the approximate courses of the central sulci and the sylvian fissures, deduced from the evoked responses to sensory stimulation. The colour codes refer to the onset latency of activation in a cortical area. Right, spatiotemporal activation sequence for individual subjects, superimposed on the schematic head. The areas of the symbols are proportional to the maximum activity (in nanoamperes, nAm) at each cortical site; the largest sphere (in S3) corresponds to 52 nAm. A minimum activation level of 10 nAm was required for a source area to be displayed; 5 nAm was sufficient if no other sources were active in the immediate temporal and spatial vicinity. Background noise level corresponds to a source intensity of 1.5 nAm.

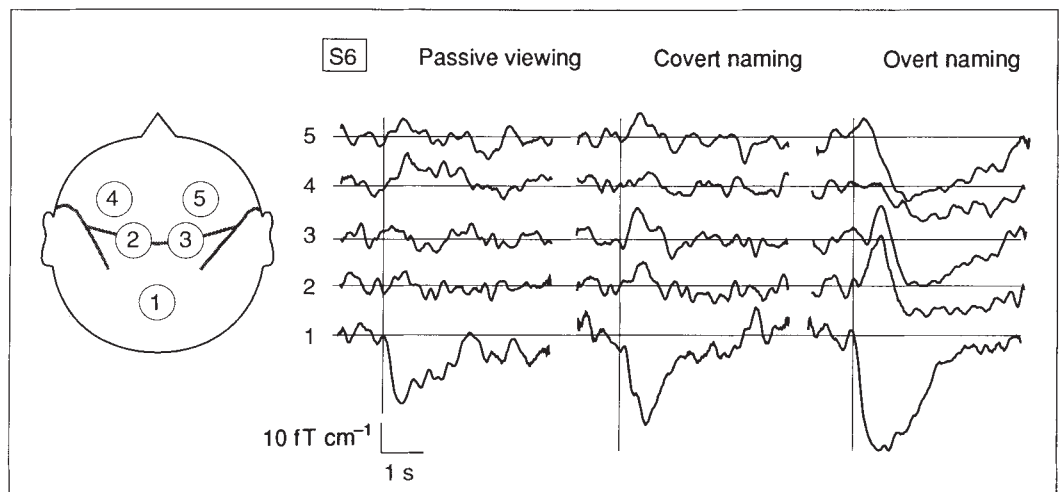
activity during the three experimental conditions (Fig. 4). The suppression of rhythmic activity, often interpreted as the involvement of cortical areas in information processing, also advanced from posterior to anterior sites. Overt naming resulted in massive reduction of rhythmic activity in the frontal regions.

Although we cannot rule out covert naming during the passive

viewing task, activation of several naming-related brain areas during this condition may indicate that humans automatically label the world around them: a failure in retrieval of symbolic interpretation could be the signal to orientate and verify.

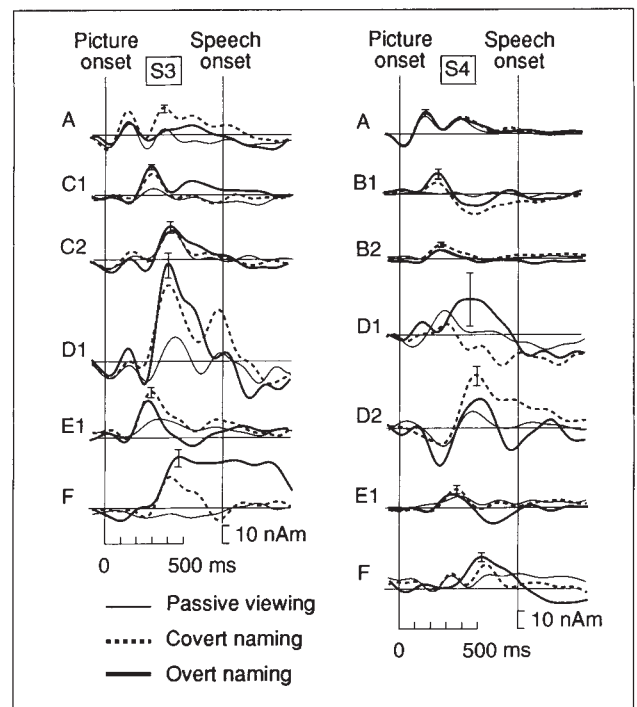
The dissimilar reactions of areas C1 and D1 to naming suggest different functions: C1 is probably a naming-specific visual area,

FIG. 4 Amplitude of 9–13-Hz spontaneous cortical activity (in femtoTesla per cm) of subject S6 as a function of time over five cortical sites (left) during the three experimental conditions. Base levels are indicated by the thin horizontal lines. Signals below and above the base level represent suppression and enhancement of the spontaneous activity, respectively. During passive viewing, only the occipital areas showed reduction of rhythmic activity. The suppression was intensified by covert naming. Overt naming led to a long-lasting effect in the occipital area, followed by suppression in the frontal sites and, finally, close to the somatomotor areas. Interestingly, while the occipital and frontal areas were involved in transforming the visual input into its spoken symbolic representation, sites 2 and 3 showed temporary enhancement of spontaneous activity, as if indicating a complete immobility.



METHODS. The data were digitally filtered through a 9–13-Hz passband. The absolute amplitude of the signal was averaged with respect to stimulus onset. This temporal-spectral-evolution method⁹ extracts features that are related to the event but not necessarily time-locked to the stimulus.

FIG. 3 Time variation of source strengths for subjects S3 (left) and S4 (right) during the three experimental conditions. The vertical bars around the maximum amplitude indicate the 95% confidence limits for source intensity. Codes A–F refer to the cortical sites shown on the left in Fig. 2. Speech onsets at 750 ms (S3) and 775 ms (S4) are depicted by vertical lines. The activations typically lasted for 200–500 ms around the peak latency. This is due partly to variance in response timing between different epochs. When speech onset was used as zero time for averaging, the temporal span of signals from area D was reduced by ~200 ms with respect to averaging performed according to the onset of visual stimulus.



whereas D1 overlaps Broca's area. Involvement of C1 in naming tasks has not been reported previously.

According to the Wernicke–Geschwind model⁸, signal analysis proceeds serially from the occipital visual areas to Broca's area via the angular gyrus and Wernicke's area. Parallel processing has recently been emphasized on the basis of positron emission tomography studies¹, which also suggested that the angular gyrus and Wernicke's area need not be involved in visual word processing. Data from subjects S4–S6 agreed with the Wernicke–Geschwind model, showing early responses from area B1. However, subjects S1–S3 showed signals in area C1 instead, possibly reflecting a conversion into symbolic form directly from the visual representation.

Tools now exist to elucidate the neural basis of human cognitive functions with a good spatiotemporal resolution. Our results on both evoked responses and reactivity of spontaneous cortical rhythms imply that signal processing associated with picture naming advances from posterior to anterior brain areas and involves both hemispheres. □

Received 26 October 1993; accepted 1 February 1994.

1. Penfield, W. & Roberts, C. *Speech and Brain Mechanisms* (Princeton Univ. Press, New Jersey, 1959).
2. Ojemann, G. A. *Semin. Neurosci.* **2**, 297–305 (1990).
3. Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. *J. cogn. Neurosci.* **1**, 153–170 (1989).
4. Hinko, R. M. *et al. Neuro Report* **4**, 675–678 (1993).
5. Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. *Rev. mod. Phys.* **65**, 413–497 (1993).

6. Ahonen, A. I. *et al.* in *Proc. Satell. Symp. Neurosci. Technol. 14th int. Conf. IEEE Engng Med. Biol. Soc.* (eds Dittmar, A. & Froment, J. C.) 16–20 (Lyon, 1992).
7. Hari, R. *J. clin. Neurophysiol.* **8**, 157–169 (1991).
8. Geschwind, N. *Cortex* **3**, 97–112 (1967).
9. Salmelin, R. & Hari, R. *Neurosci.* (in the press).

ACKNOWLEDGEMENTS. This study was supported by the Academy of Finland and the Sigrid Jusélius Foundation. The MRI data were obtained at the Department of Radiology of Helsinki University Hospital. We thank S. Williamson and M. Laine for comments on the manuscript.

An RNA polymerase II holoenzyme responsive to activators

Anthony J. Koleske & Richard A. Young

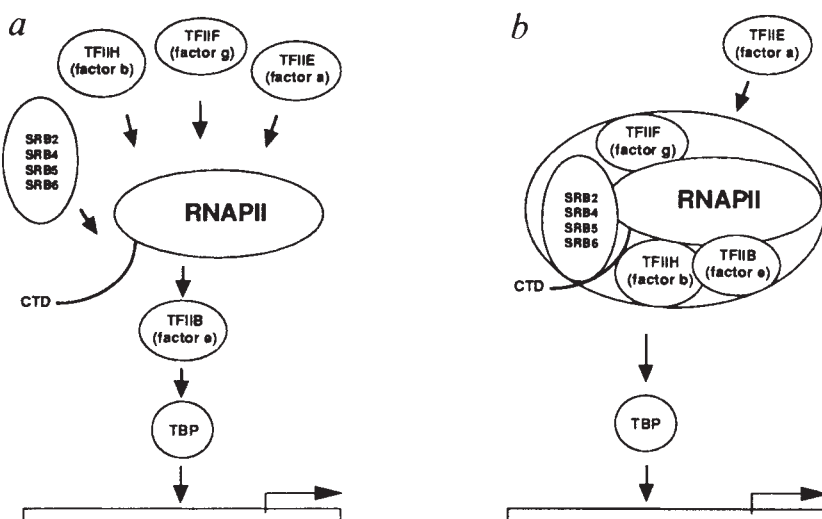
Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

RNA POLYMERASE II requires multiple general transcription factors to initiate site-specific transcription¹⁻³. These proteins can assemble in an ordered fashion onto promoter DNA *in vitro*²⁻⁸, and such ordered assembly may occur *in vivo* (Fig. 1a). Some general transcription factors can interact with RNA polymerase II in the absence of DNA^{3,9-15}, however, suggesting that RNA polymerase II may also assemble into a multi-component complex containing a subset of initiation factors before binding to promoter DNA (Fig. 1b). Here we present evidence from the yeast *Saccharomyces cerevisiae* for such an RNA polymerase II holoenzyme, a multi-subunit complex containing roughly equimolar amounts of RNA polymerase II, a subset of general transcription factors, and SRB regulatory proteins. Transcription by this holoenzyme is stimulated by the activator protein GAL4-VP16, a feature not observed with purified RNA polymerase II and general transcription factors alone. We propose that the holoenzyme is a form of RNA polymerase II readily recruited to promoters *in vivo*.

The yeast SRB proteins, which play essential roles in transcription initiation *in vivo* and *in vitro*, copurify with RNA

polymerase II and additional unidentified polypeptides in a high M_r complex^{16,17}. To clarify the role of the complex in transcriptional initiation, we searched for additional components needed for selective transcription *in vitro* (Fig. 2). Because the complex contains equimolar amounts of RNA polymerase II and SRB protein molecules, but submolar amounts of TATA-binding protein (TBP) (see below), TBP levels needed to be supplemented with recombinant TBP. Specific transcription of promoter-containing DNA also required a fraction from a yeast whole-cell extract. Purification of this activity revealed that it is composed of two polypeptides (Fig. 2b) whose chromatographic behaviour and size (66K and 43K) are identical to that for factor a¹⁸, the yeast homologue of TFIIE (Fig. 1c). The RNA polymerase II-containing complex thus represents a novel form of the enzyme which, unlike purified RNA polymerase II¹⁹, requires only TBP and factor a for site-specific initiation (Fig. 2c). These results indicated that it might contain some of the general transcription factors pre-assembled to form an RNA polymerase II holoenzyme.

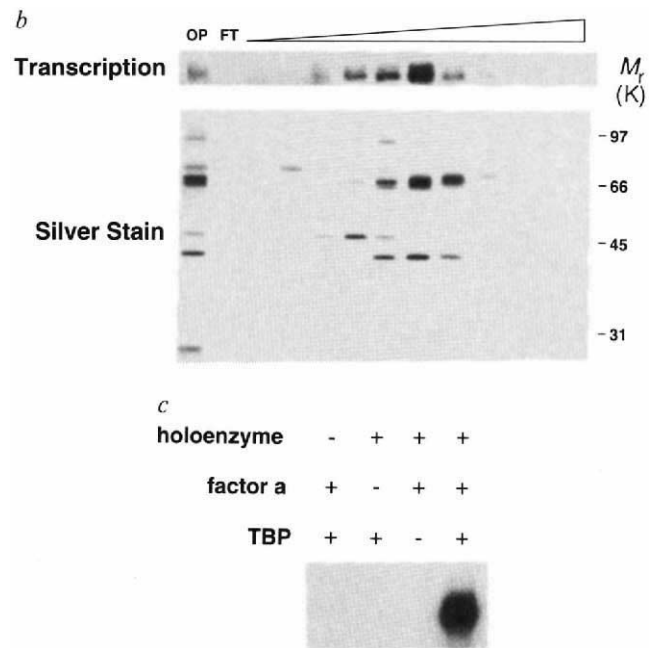
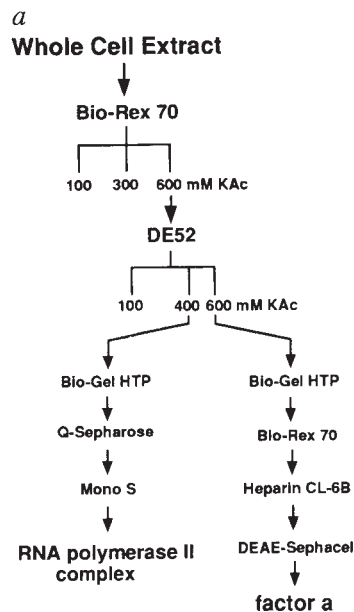
Five general factors (a, b, d, e and g) are sufficient to allow yeast RNA polymerase II to initiate transcription accurately *in vitro*¹⁹. To determine which of these were associated with the complex, column fractions from the final purification step were tested in reconstituted transcription reactions and analysed by western blotting with antisera to yeast initiation factors (Fig. 3a). Transcriptional activity coeluted with RNA polymerase II and SRB2, 4, 5 and 6. The activity also coeluted with the 41K yeast factor e (TFIIB) protein^{15,20} and the 73K TFB1 subunit of yeast factor b (TFIIH)²¹. Specific antisera are not yet available for factor g (TFIIF), but the purified complex (Fig. 3b) contains three polypeptides of the same size as the subunits of purified factor g (105K, 55K and 30K)¹⁴. Human RNA polymerase II



Yeast	Rat	Human
d (TBP)	τ (native TATA factor)	TFIID
e (SUA7)	α	TFIIB
b	δ	TFIIH (BTF2)
g	$\beta\gamma$	TFIIF (RAP30/74)
a	ϵ	TFIIE
TFIIA		TFIIA

FIG. 1 Models for RNA polymerase II transcription initiation complex formation. a, Stepwise association of RNA polymerase II and transcription factors into an initiation complex. In this model, each of the initiation factors and RNA polymerase II assembles onto the DNA in a stepwise manner. b, Assembly of an RNA polymerase II holoenzyme into an initiation complex. In this model, a subset of initiation factors can assemble with RNA polymerase II into a holoenzyme in the absence of DNA. This holoenzyme and factor a associate with a complex of TATA-binding protein (TBP) and DNA to form an initiation complex. The two models are not mutually exclusive. c, Yeast RNA polymerase II initiation factors and homologues in other species (ref. 3 and W. J. Feaver, D. Bushnell, N. L. Henry and R. D. Kornberg, personal communication).

FIG. 2 The RNA polymerase II holoenzyme, TBP, and factor a are sufficient for transcription initiation *in vitro*. **a**, The RNA polymerase II holoenzyme (RNA polymerase II complex) and factor a were purified according to this chromatographic scheme^{16,18}. **b**, Factor a is required in addition to TBP and the RNA polymerase II complex for *in vitro* transcription. Semi-purified factor a (300 µg protein in 2 ml) eluted from the heparin-CL-6B column was loaded onto a DEAE-Sephacel column and eluted with a 0.15–1.0 M gradient of potassium acetate. The output (OP) and flow-through (FT) and a portion of every other fraction eluting from this column between 0.32 and 1.0 M potassium acetate were analysed for transcriptional activity (top) and for the presence of polypeptides by SDS-PAGE (bottom). Top, assays were done using pGAL4G-template²⁹ (300 ng), RNA polymerase II complex (1 µg), recombinant TBP (40 ng), and 1 µl of the OP, FT, and every other fraction from the column^{16,17,30}. Bottom, 2.5 µl of the OP, FT and every other column fraction was subjected to electrophoresis on a 12% SDS-PAGE gel^{16,17}, and the gel was stained with silver using standard protocols^{16,17}. The positions of M_r standards are shown on the right of



the panel. **c**, The holoenzyme, factor a and TBP are sufficient for *in vitro* transcription. Transcription reactions were done using the pGAL4G-template (300 ng) and standard conditions^{4,6,17,30}. The holoenzyme (1 µg), factor a (40 ng) and recombinant TBP (40 ng) were added to reactions as indicated. This and other figures in this paper were prepared from digital replicas of primary data scanned using a UMax UC840 Max Vision digital scanner.

requires TFIIF and TFIIF to transcribe linear templates²², and the fact that the yeast holoenzyme can also transcribe such templates (see below) indicates that it indeed contains TFIIF and TFIIF activity. The purified complex thus represents a form of RNA polymerase II that is tightly associated with multiple SRB proteins and with factors b, e and g (TFIIF, TFIIB and TFIIF), and can accurately initiate transcription when supplemented with factor a (TFIIF) and TBP.

The holoenzyme is highly stable; it remains intact on chromatography through six ion exchange columns, and migrates as a single 1,200K complex on gel filtration¹⁶. To confirm that it is a single multi-subunit complex, immunoprecipitation experiments were done. The four SRB proteins, factor e (TFIIB), the TFIIB subunit of factor b (TFIIF), and RNA polymerase II all specifically precipitated from purified preparations of the holoenzyme by anti-SRB5 antibodies (Fig. 3c). Similar results were obtained when the complex was precipitated with antibodies against other holoenzyme components.

The holoenzyme preparation contains roughly equimolar amounts of SRB2, SRB5, factor e (TFIIB) and RNA polymerase II (Fig. 3d, e), but does not contain significant amounts of the TOA1 subunit²³ of yeast TFIIF (Fig. 3d). Although we previously reported that it contained some TBP, the purest form of the holoenzyme contains less than one molecule of TBP per 50 molecules of RNA polymerase II, and its transcriptional activity is absolutely dependent on the addition of recombinant TBP (Fig. 2c). At each step in the purification, some TBP coelutes from the column with the holoenzyme while some elutes as free TBP (ref. 16 and our unpublished observations). This behaviour may reflect a weak interaction between the holoenzyme and TBP in the absence of DNA, as the purified holoenzyme contains no detectable DNA (not shown). TBP can bind to SRB2 (ref. 17), SRB5 (ref. 16), and the RNA polymerase II carboxy-terminal

domain (CTD)^{16,24} on affinity columns, suggesting that it may interact with these components of the holoenzyme.

Purified yeast RNA polymerase II is not stimulated by transcriptional activators in the presence of general transcription factors alone^{25,26}. Transcription of supercoiled or linear templates by the holoenzyme, TBP and factor a, however, was stimulated approximately fivefold by the transcriptional activator GAL4-VP16 (Fig. 4a, b). For comparison, transcription of a supercoiled template in crude yeast nuclear extracts was stimulated 10-fold under the same conditions (Fig. 4a). One or more components of the holoenzyme can thus respond to activation signals from GAL4-VP16.

Others have postulated that RNA polymerase II enters the initiation complex as a holoenzyme containing a subset of the general initiation factors, because some general factors (including TFIIB, TFIIF, TFIIF) can associate with RNA polymerase II in the absence of DNA^{3,9-15}. The holoenzyme described here is tightly associated with these general initiation factors.

The holoenzyme probably escaped earlier detection because it is much less abundant than free RNA polymerase II. Although the holoenzyme contains most of the SRB protein in whole cell extracts, it only contains 6% of RNA polymerase II and 12% of TFIIB (Fig. 3e). We estimate that yeast haploid cells contain roughly 1,000 molecules of the holoenzyme, a number adequate to initiate transcription at active promoters. The proportion of active promoters transcribed by the holoenzyme is not yet known. A significant fraction of cellular RNA polymerase II is involved in elongation of nascent transcripts, accounting for at least some of the enzyme not complexed with SRB proteins.

The existence of the holoenzyme has implications for the mechanisms involved in the regulation of transcription. Activators appear to function, at least in part, through interactions with multi-subunit TFIID, and the holoenzyme may be recruited

FIG. 3 The holoenzyme is a complex of RNA polymerase II and initiation factors. *a*, Semipurified holoenzyme that eluted from the Q-Sepharose column (Fig. 2a) was loaded onto a Mono-S column and eluted with a 0.1–1.0 M gradient of potassium acetate. The output (OP) and flow-through (FT) and a portion of every other fraction eluting between 0.1 and 0.9 M potassium acetate were analysed for holoenzyme activity (top). These samples were also analysed by western blot for the presence of RNA polymerase II and transcription factors (bottom). Top, Transcription assays used the pGAL4G- template (300 ng), factor a (40 ng), recombinant TBP (40 ng) and 1 µl of the OP, FT and every other fraction from the Mono-S column. Bottom, 1 µl of the same fractions were also separated on an SDS–polyacrylamide gel and blotted to nitrocellulose^{16,17}. The blots were probed with polyclonal antibodies specific to the 73K subunit of factor b (TFB1), factor e (SUA7), SRB2, SRB4, SRB5 and SRB6 and monoclonal antibodies specific to the largest subunit of RNA polymerase II (RPB1)^{15,16,19,20}. *b*, Polypeptide composition of RNA polymerase II holoenzyme. Purified holoenzyme (1 µg) was subjected to SDS–PAGE and stained with silver^{16,17}. Western blots of purified holoenzyme were done on samples run on adjacent lanes of the gel with antiserum used in *a* to identify subunits of the SRB complex. Proteins in the holoenzyme preparation that correspond in size to subunits of RNA polymerase II, SRB proteins, or subunits of initiation factors are indicated. The sizes of protein *M_r* standards are indicated in K ($\times 10^3$). *c*, Coimmunoprecipitation of holoenzyme components with SRB5. Purified RNA polymerase II (15 µg) was diluted in 0.5 ml of transcription buffer containing potassium acetate instead of potassium glutamate, 0.01% NP-40 and 0.1 mg ml⁻¹ BSA. Affinity-purified anti-HSP70 or anti-SRB5 antibodies (1 µg), and 5 µg recombinant SRB2 or SRB5 protein were added as indicated. Immunoprecipitated material was analysed by western blot as indicated in *b* for the presence of transcription factor subunits and RNA polymerase II. *d*, Quantification of holoenzyme components. Samples of whole cell extract, nuclear extract and purified holoenzyme together with standard amounts of purified RNA polymerase II and recombinant transcription factor subunits were quantified by western blotting. Each gel contained the following: 25 µg yeast whole cell extract (lane 1), 25 µg yeast nuclear extract (lane 2), 1 µg purified holoenzyme (lane 3) and 0.2 µg purified holoenzyme (lane 4). The gels also contained purified standard proteins in lanes 5–7 in the following amounts: 8, 40 and 200 ng RNA polymerase II; 4, 20 and 100 ng 6His-tagged factor e (SUA7); 3.2, 16 and 80 ng SRB2; 4, 20 and 100 ng SRB5; 3.2, 16 and 80 ng TOA1; 3.2, 16 and 80 ng TBP. Epitope-tagged SRB2 and 6His-tagged factor e (SUA7) used in this analysis exhibit slightly lower mobility on gels than their untagged counterparts. The RNA polymerase II CTD in the holoenzyme is the hypophosphorylated form (II_A). *e*, Summary of holoenzyme components. The amount of each holoenzyme component in 1 µg of the holoenzyme was determined by comparison with standard amounts in *d*. Taking the *M_r* of each component into account, the stoichiometry of each factor per RNA polymerase II molecule is presented.

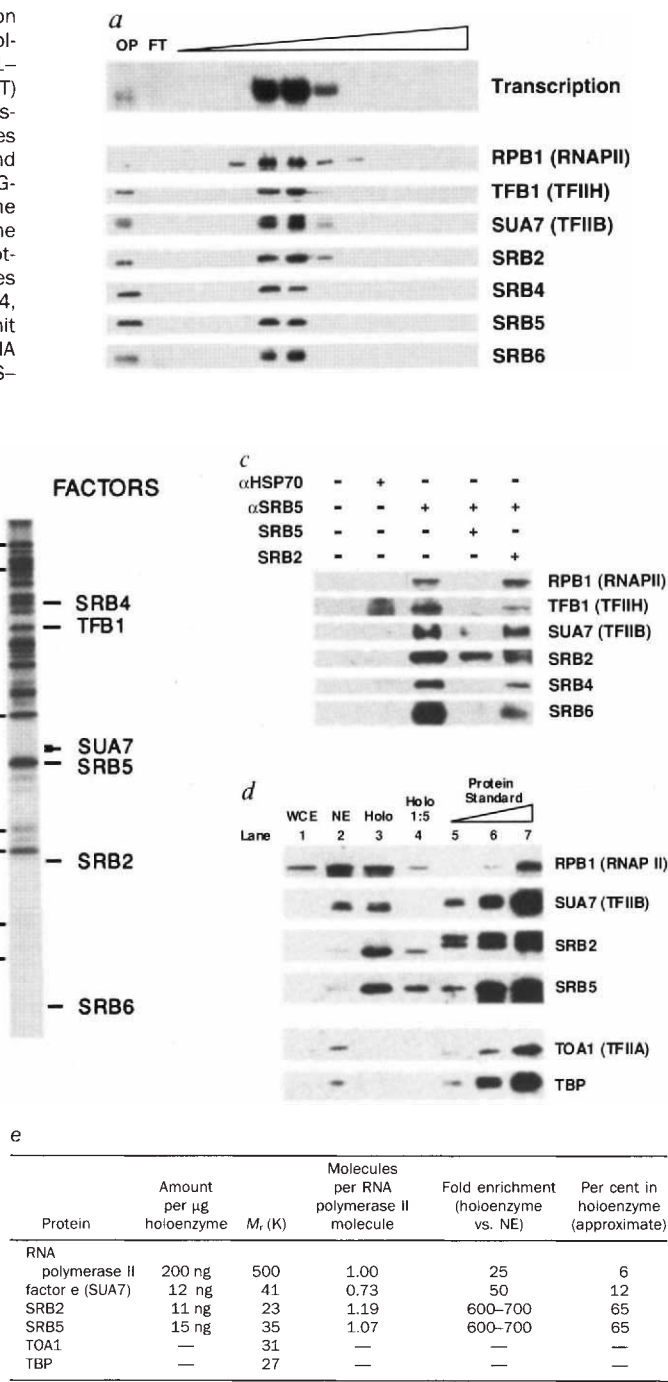


FIG. 4 Transcription by the holoenzyme is stimulated by GAL4-VP16. *a*, Transcription reactions were done using either a template containing a CYC1 TATA element that directs transcription of a G-less cassette (GAL4 site template) or a template that contained in addition a single consensus DNA-binding site for the GAL4 protein upstream of the TATA element (+GAL4 site template). GAL4-VP16 (150 ng) was added to reactions as indicated. Top, Reactions were done with the holoenzyme (1 µg), factor a (40 ng), recombinant TBP (40 ng), and each template (100 ng) as indicated. Bottom, Reactions were done with yeast nuclear extract protein (150 µg). Transcription in reactions containing nuclear extract is stimulated 10-fold by GAL4-VP16. Transcription by the holoenzyme is stimulated fivefold by GAL4-VP16. +GAL4 site template is pGAL4G– (ref. 29). –GAL4 site template is pSL187 (ref. 17). The exposure in the top panel was 5 times longer than the exposure in the bottom panel. Levels of transcript were quantified using a Fuji Bio-image Analyzer. *b*, Reactions were done with the holoenzyme as detailed above except 225 ng of template, linearized by digestion with *Pvu*II restriction endonuclease, was used. This exposure was 3 times longer than the holoenzyme panel in *a*.

to promoters through interactions with activators and promoter-bound TFIID. The activation observed in crude extracts is double the activation obtained with the purified holoenzyme, which may reflect the absence of TAFs in the holoenzyme reactions.

The SRB proteins appear to act as a regulatory 'glue' that stabilizes interactions between RNA polymerase II and general

transcription factors. They may also confer responsiveness to transcriptional activators. In addition, some SRB proteins may regulate events subsequent to initiation complex formation, such as phosphorylation of the CTD^{27,28} or promoter clearance. Additional insights into the mechanisms involved in the regulation of initiation should be obtained through further study of SRB proteins and the holoenzyme. □

Received 31 December 1993; accepted 15 February 1994.

1. Sawadogo, M. & Sentenac, A. A. *Rev. Biochem.* **59**, 711–754 (1990).
2. Zewel, L. & Reinberg, D. *Curr. Opin. Cell Biol.* **4**, 488–495 (1992).
3. Conaway, R. C. & Conaway, J. W. A. *Rev. Biochem.* **62**, 161–190 (1993).
4. Davison, B. L., Egly, J. M., Mulvihill, E. R. & Chambon, P. *Nature* **301**, 680–686 (1983).
5. Fire, A., Samuels, M. & Sharp, P. A. *J. biol. Chem.* **259**, 2509–2516 (1984).
6. Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. *Science* **241**, 1335–1338 (1988).
7. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **58**, 549–561 (1989).
8. Flores, O., Lu, H. & Reinberg, D. *J. biol. Chem.* **267**, 2786–2793 (1992).
9. Sopta, M., Carthew, R. W. & Greenleaf, A. L. *Molec. cell. Biol.* **9**, 1465–1475 (1989).
10. Price, D. H., Sluder, A. E. & Greenleaf, A. L. *Molec. cell. Biol.* **9**, 1465–1475 (1989).
11. Flores, O., Ha, I. & Reinberg, D. *J. biol. Chem.* **265**, 5629–5634 (1990).
12. Kitajima, S. et al. *Nucleic Acids Res.* **18**, 4843–4849 (1990).
13. Gerard, M. et al. *J. biol. Chem.* **266**, 20940–20945 (1991).
14. Henry, N. L., Sayre, M. H. & Kornberg, R. D. *J. biol. Chem.* **267**, 23388–23392 (1992).
15. Tschochner, H., Sayre, M. H., Flanagan, P. M., Feaver, W. J. & Kornberg, R. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11292–11296 (1992).
16. Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. *Cell* **73**, 1367–1375 (1993).
17. Koleske, A. J., Buratowski, S., Nonet, M. & Young, R. A. *Cell* **69**, 883–894 (1992).
18. Sayre, M. H., Tschochner, H. & Kornberg, R. D. *J. biol. Chem.* **267**, 23383–23387 (1992).
19. Sayre, M. H., Tschochner, H. & Kornberg, R. D. *J. biol. Chem.* **267**, 23376–23382 (1992).

20. Pinto, I., Ware, D. E. & Hampsey, M. *Cell* **68**, 977–988 (1992).
21. Gileadi, O., Feaver, W. J. & Kornberg, R. D. *Science* **267**, 1389–1391 (1992).
22. Parvin, J. D. & Sharp, P. A. *Cell* **73**, 533–540 (1993).
23. Ranish, J. A., Lane, W. S. & Hahn, S. *Science* **255**, 1127–1132 (1992).
24. Usheva, A. et al. *Cell* **69**, 871–881 (1993).
25. Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H. & Kornberg, R. D. *Nature* **350**, 436–438 (1991).
26. Flanagan, P. M., Kelleher, R. J., Tschochner, H., Sayre, M. H. & Kornberg, R. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7659–7663 (1992).
27. Payne, J. M., Laybourn, P. J. & Dahmus, M. E. *J. biol. Chem.* **264**, 19621–19629 (1989).
28. Laybourn, P. J. & Dahmus, M. E. *J. biol. Chem.* **265**, 13165–13173 (1990).
29. Lue, N. F., Flanagan, P. M., Sugimoto, K. & Kornberg, R. D. *Science* **246**, 661–664 (1989).
30. Liao, S.-M., Taylor, I. C. A., Kingston, R. E. & Young, R. A. *Genes Dev.* **5**, 2431–2440 (1991).

ACKNOWLEDGEMENTS. We thank J. Feaver, Y.-J. Kim and R. Kornberg for anti-TFB1 antiserum and affinity-purified anti-SRB5 antibodies, J. Ranish and S. Hahn for TOA1-expressing bacterial clones and anti-TOA1 antiserum, L. Strasheim and R. Burgess for purified yeast RNA polymerase II, R. Knaus and L. Guarente for recombinant 6 His-SUA7 and anti-SUA7 antiserum, D. Jeffery and P. Murray for affinity-purified anti-HSP70 antibodies, and C. Thompson for GAL4-VP16 protein and assistance in growing the yeast. We thank S. Buratowski, D. Chao, L. Guarente, P. Guilfoile, S. Hahn, M. Hampsey, D. Jeffery, R. Kornberg, Y. Li, S.-M. Liao, P. Murray, M. Sayre, P. Sharp, C. Thompson and F. Winston for comments, criticisms and discussions.

Determinants of repressor/operator recognition from the structure of the *trp* operator binding site

Z. Shakked*, G. Guzikevich-Guerstein*, F. Frolow*, D. Rabinovich*, A. Joachimiak† & P. B. Sigler†

* Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

† Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06510, USA

ON the basis of the crystal structure of the *trp* repressor/operator complex¹, it has been proposed that the specificity of the interaction can be explained not only by direct hydrogen bonding and non-polar contacts between the protein and the bases of its target DNA, but also by indirect structural effects and water-mediated interactions. To understand the contribution of DNA structure and hydration in this context, the structure of the free DNA must be compared with its structure when complexed with the protein. Here we present the high-resolution crystal structure of the *trp* operator region that is most important in the recognition process. By comparing the free and bound states of the DNA regulatory sequence, we show that the structure and hydration of the DNA target are important elements in its recognition by the repressor protein.

We have investigated the crystal structures of the 18-base-pair (bp) *trp* operator region TGTACTAGTTAACTAGTAC and of shorter fragments incorporating the 6-bp binding site ACTAGT. We now report the crystal structure of the B-form DNA decamer CCACTAGTGG determined at 1.95 Å resolution. In the crystal, the B-type duplexes are extensively hydrated and stacked end-to-end to form continuous polymers (Fig. 1a). The other short intermolecular contacts where the backbone of one duplex interacts with the major or the minor groove of a neighbouring duplex are shown in Fig. 1b. This crystal form is less tightly packed than other forms and therefore imposes fewer packing constraints on the helices that adopt it^{2,3}.

A comparison between the uncomplexed 6-bp recognition site in the decamer and the corresponding 6-bp regions in the complex is shown in Fig. 1c. Whereas this segment in the decamer is straight, the one in the complex is bent by 15° in a direction that results in a compression of the major groove and an expansion of the minor groove in the bound fragment with respect to the free DNA. Despite such differences, there is a striking simi-

TABLE 1 Base-stacking parameters of the free decamer (f) and the complexed operator (c)

Helical step	Helix twist (°)		Roll (°)		Slide (Å)	
	f	c	f	c	f	c
C-C G-T	29.1	25.9	1.5	-0.3	0.2	-0.4
C-A T-A	38.9	42.8	5.8	-0.6	0.5	-0.2
A-C	27.6	30.5	-1.3	-0.7	-0.8	-1.1
C-T	36.3	32.5	-1.1	4.6	-0.6	-0.9
T-A	38.3	31.2	1.0	11.7	-0.7	-1.0
A-G	37.7	36.9	4.8	4.7	-0.3	-0.2
G-T	30.8	34.1	-3.9	-0.2	-0.6	-0.9
T-G T-T	32.2	38.4	3.9	1.4	1.1	0.3
G-G T-A	34.4	31.7	1.4	6.0	0.7	0.9
Average [*]	33.9	33.8	1.3	3.0	-0.1	-0.4
Average [†]	34.1	33.0	-0.1	4.0	-0.6	-0.8

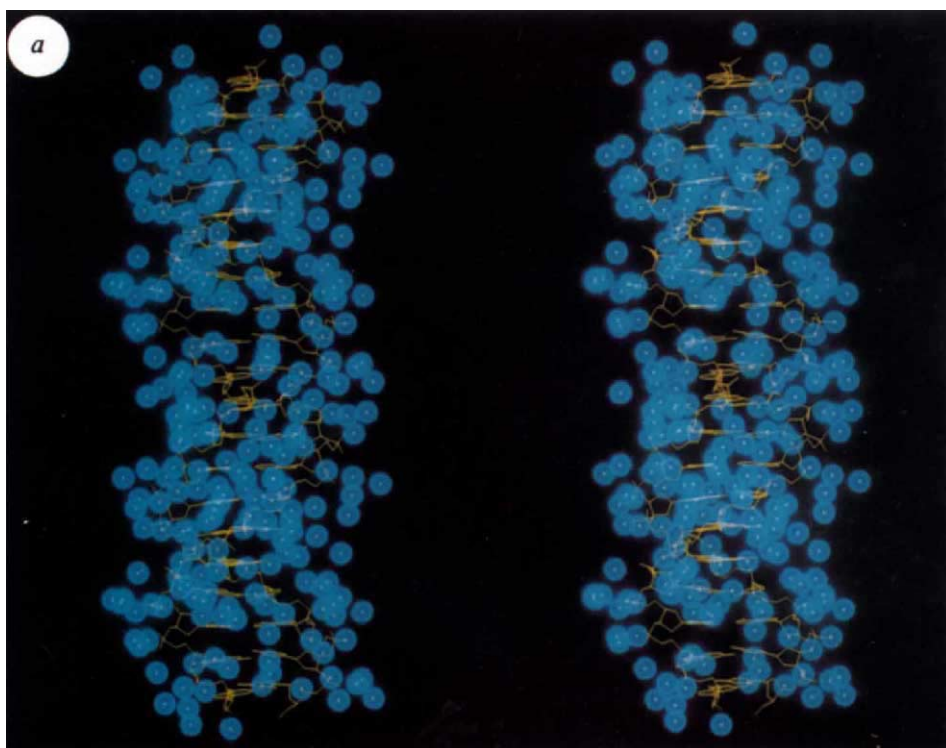
The stacking geometry of two adjacent base pairs can be defined by their relative rotations and translations¹⁴. Two rotations (helix twist and roll) and one translation (slide) are significantly variable in DNA crystal structures⁹. Twist is the relative rotation of the two base pairs about the helix axis. The roll is the angular opening or closing of the base pairs towards the helical grooves. The slide is the relative displacement of the base pairs along their long axes. Only the bases of the top strand are indicated for each helical step. The numbering scheme is as for Fig. 2. Hence, the first step of the free target (C-C) consists of base pairs C₋₉ · G₉ and C₋₈ · G₈, and that of the complexed target (G-T) consists of G₋₉ · C₉ and T₋₈ · A₈, the central five steps that are common to both sequences are highlighted in bold. Calculations were done with NEWHEL93 by R. E. Dickerson. Data for the complexed operator are based on the high-resolution refinement (1.9 Å) of the complex¹¹; stacking parameters were obtained by averaging equivalent values of the four crystallographically independent half-sites. The largest discrepancy between free and bound DNA is shown by the central T-A step, as reflected by the corresponding helix twist and roll angles. The bending of the bound helix is achieved primarily by the positive rolling of successive base pairs in the central region (4.6, 11.7, 4.7°); that is, opening towards the minor groove and closing towards the major groove. The corresponding roll values of the straight decamer are small. The negative slide values of the central regions, like base-pair displacements (Fig. 2a), are indicative of a deep major groove. The sign convention follows ref. 14 and is reversed from that published previously for slide values¹.

* All steps.

† Central five steps.

FIG. 1 a, Stereo drawing of the DNA decamer in the crystal structure. The 20-bp helix (yellow) and the surrounding water molecules, mainly from the first and second hydration shells (blue spheres), were generated by the crystal's symmetry. b, Intermolecular interactions. Stereo drawing of two continuous helices related by a crystallographic two-fold axis and tilted relative to each other by 120° . Bases are shown in purple and the sugar-phosphate backbone is yellow for one helix and light blue for the other. The two polymers interact through the major groove and the minor groove of one helix and the sugar-phosphate backbone of another. A detailed description of the packing arrangement of this trigonal form based on the two other structures which have similar packing to this is given in refs 2 and 3. A similar type of major groove/backbone interaction was first observed in a B-DNA dodecamer²⁰. c, Superposition of the target's ACTAGT region (green) of the decamer onto the equivalent half-sites of the complexed operator (red). It shows that the 6-bp region in the complex is bent towards the major groove, whereas that of the free decamer is straight. Thus the major groove of the complexed site is narrower and its minor groove is wider than that of the decamer. The widths of the major and minor grooves of the complexed half-site are 16 and 13 Å; the equivalent values for the free site are 19 and 10 Å. These values were obtained by averaging the shortest P-P distances across each 6-bp groove. The r.m.s. deviations between the free half-site and each of the four unique ones in the complex structure range from 1.2 to 1.4 Å. The comparison is based on the high-resolution refinement (1.9 Å) of the complex¹¹.

METHODS. The self-complementary oligonucleotide d(CCACTAGTGG)



was synthesized by the phosphoramidite method and purified by reverse-phase HPLC and then by ion-exchange chromatography. Crystals were grown at 4°C over a period of 4 weeks from 5- μl drops containing 2 mg ml^{-1} DNA, 20 mM cacodylate buffer, pH 7.0, 4 mM spermine-4HCl, 30 mM MgCl_2 and 10% polyethyleneglycol 200 (PEG200) and equilibrated against aqueous solutions containing 36% PEG200 and 100 mM cacodylate, pH 7. Before data collection, crystals

larity in several features of helical geometry. Both the 19-mer in the complex and the free decamer are significantly underwound (Table 1). The mean helix twist is 34° , corresponding to 10.6 bp per turn for both fragments, and not the 10 bp per turn found in most high-resolution B-form DNA duplexes⁴⁻⁸. Both the bound and free DNA fragments have unusually deep major grooves. The variations in depth of the major groove can be described by the distances of the base pairs from the helix axis (Fig. 2a). This parameter (termed X-displacement) is important for distinguishing between A- and B-form DNA configurations⁹. In the complexed operator and uncomplexed decamer, the mean values are negative (-1.9 and -0.9 Å, respectively), implying that the base pairs have been displaced from the helix axis to create a deep major groove. The corresponding averages of the hexameric recognition regions are even more negative (-2.0 and -1.4 Å; Fig. 2a). The values reported for most crystalline B-form DNAs range from 0 to 1 Å (refs 4-8).

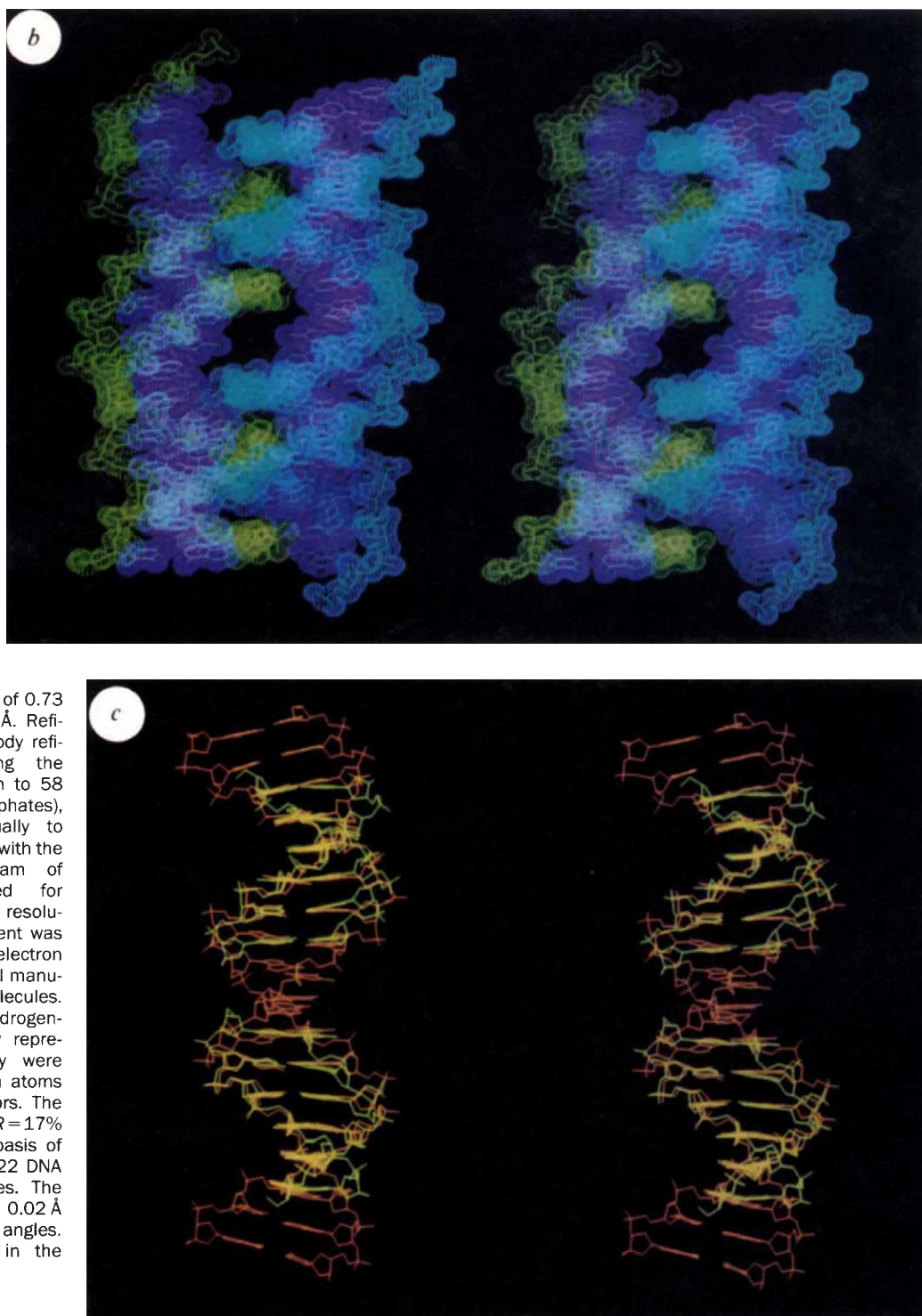
Although the backbone parameters of the recognition sites in the complex are typical of a B-form DNA, the disposition of the base pairs with respect to the helix axis is midway between an A- and a B-type helix (Fig. 2a and b). The structure of the free target sequence deviates significantly from that of other crystalline B-DNAs and is more like its structure in the complex. Also, the contour of the deep major groove of the free fragment mimics that of the complex (as the corresponding base pairs are displaced to the same degree relative to their neighbours), although the groove is deeper in the complex by ~ 1 Å (Fig. 2a). When the complex is formed, the DNA bends by nearly 15° at each half-site. Such deformation is accommodated by changes in the base stacking of the recognition regions (Table 1). We suggest that the energy of deformation is likely to be low because of the structural similarity between the free and complexed tar-

gets. It has been proposed that the relative instability or deformability of the central T-A step could facilitate the formation of a stable complex by providing a low-energy barrier to any necessary structural adjustments on binding to the protein¹⁰. Thus in the *trp* repressor/operator system, the characteristic structure of the specific base sequence and its inherent flexibility seem to be major determinants in the recognition process.

Another important contribution to specificity in this system comes from water-mediated contacts between the DNA and protein^{1,11}. As both crystal structures were determined at high resolution (better than 2 Å), they allow investigation into the first and second hydration shells of the DNA target in its complexed and uncomplexed forms. A detailed analysis of the hydration in the crystal structure of the free decamer and each of the four crystallographically independent half-sites in the crystal structure of the complex reveals ten conserved hydration sites in the major groove. The water molecules occupying these sites form direct hydrogen bonds with the same ten bases in the recognition sequence of the free and the bound DNA (Fig. 3a). The ten hydration sites are arranged in a pseudo-symmetrical manner, following the internal symmetry of the common 6-bp sequence in both the complexed and uncomplexed fragments. Three water molecules of the bound target mediate specific contacts between the protein and three purines (A_{-7} , G_6 and A_5) that are among the most highly conserved and mutationally sensitive bases of the operator (Fig. 3a). These three hydration sites are arranged almost identically to those that hydrate the same purines in the free decamer (Fig. 3b).

In both free and bound target, the surface of every purine in the major groove is hydrated (Fig. 3a). When the purine is singly hydrated, as it is in most instances here, the water molecule usually forms two hydrogen bonds, one to N7 and the other to

were transferred to solutions containing 60% 2-methyl-2,4-pentenediol in 50 mM cacodylate, pH 7. Data were collected on a Rigaku AFC5R diffractometer from one crystal ($0.4 \times 0.4 \times 0.1 \text{ mm}^3$) at 100 K with graphite-monochromated $\text{CuK}\alpha$ radiation generated by a rotating anode operated at 10 kW. The space group is $P3_221$ (as established by the analysis), and cell dimensions are $a=b=32.9 \text{ \AA}$ and $c=95.1 \text{ \AA}$, with one decamer duplex in the asymmetric unit. A total of 8,200 reflections to $1.95 \text{ \AA}$ resolution were measured by the ω scanning method and averaged to 4,034 unique reflections ($R_{\text{sym}}(I)=8.7\%$ for $I>0$). The structure solution was carried out by ULTIMA¹⁵ using a standard B-DNA model¹⁶ with the two possible space groups ($P3_221$ or $P3_221$). The best trial structure on the basis of R -factor and correlation coefficient was obtained for space group $P3_221$. The structure was refined as a rigid body to $R=0.39$ and a correlation coefficient of 0.73 in the resolution range of $25\text{--}6 \text{ \AA}$. Refinement continued with a rigid-body refinement by X-PLOR¹⁷, dividing the structure to 10 groups and then to 58 groups (bases, sugars and phosphates), changing the resolution gradually to $8\text{--}3 \text{ \AA}$. Refinement was continued with the restrained least-squares program of Konnert-Hendrickson¹⁸, modified for nucleic acids¹⁹, using data in the resolution range $6\text{--}1.95 \text{ \AA}$. The refinement was interspersed with examination of electron density maps to correct the model manually and to locate solvent molecules. Solvent molecules satisfying hydrogen-bonding criteria and adequately represented in the electron density were included in the model as oxygen atoms with individual temperature factors. The structure is at present refined to $R=1.7\%$ (for 3,200 observations on the basis of $F \geq 2\sigma(F)$) with the inclusion of 422 DNA atoms and 123 water molecules. The r.m.s. deviations from ideality are $0.02 \text{ \AA}$ in bond lengths and 2.5° in bond angles. Coordinates will be deposited in the Protein Data Bank.



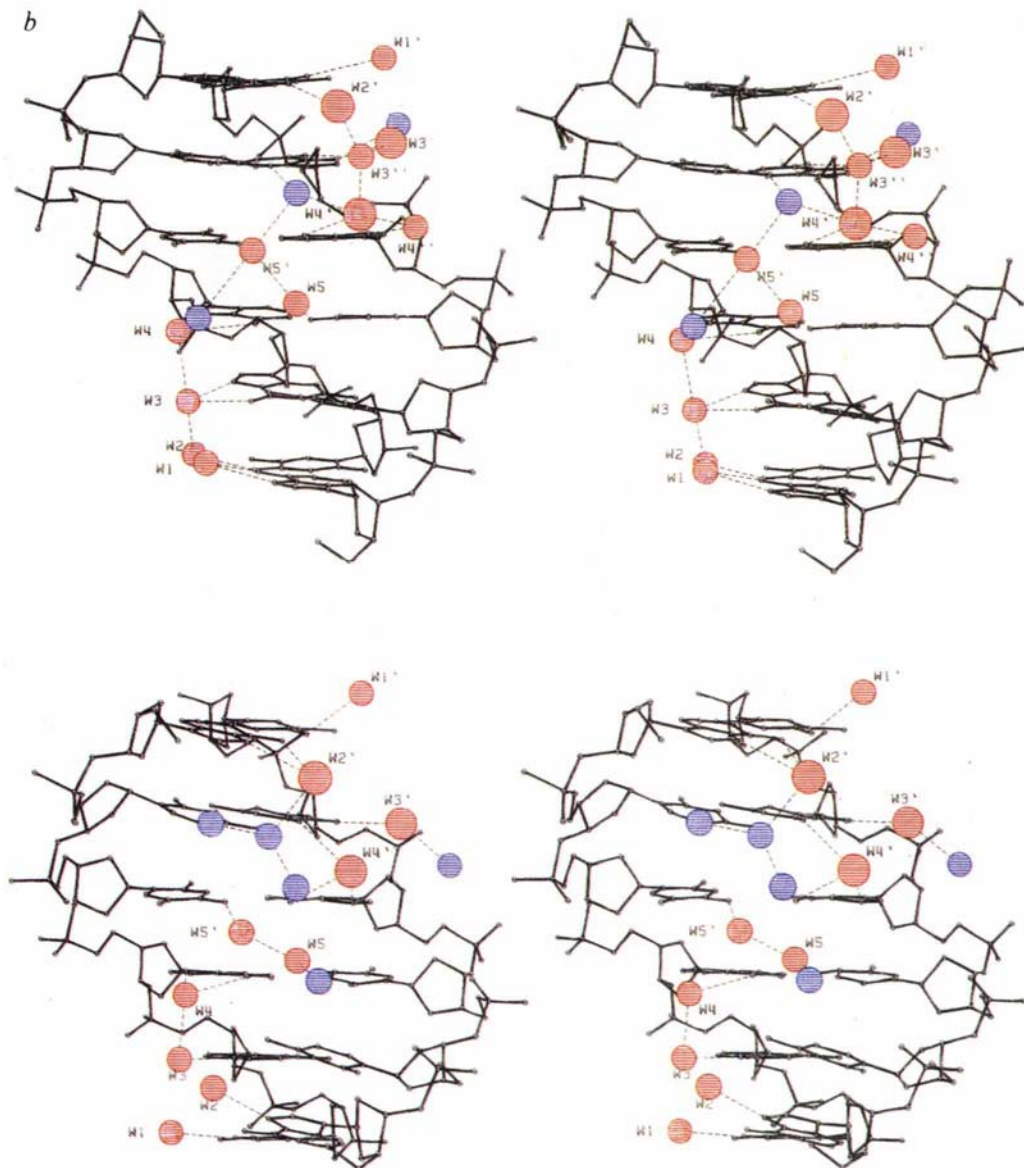
N6 or O6. In the few cases in which the purine is doubly hydrated, each of the two water molecules forms a single hydrogen bond to either of the polar atoms, as seen at A_5 and G_6 of the free decamer (Fig. 3b, c). The three critical purine bases (A_{-7} , G_6 and A_5) of the bound target are singly hydrated in all four crystallographically independent half-sites of the complex as there is no room for additional water molecules at the interface between the two macromolecules. If the higher hydrated state is encountered as the protein approaches its target, the extra solvent is expelled at an enthalpy cost that is probably compensated by a favourable entropy change and by the formation of new hydrogen bonds to the protein.

This comparative analysis of the water structure in the com-

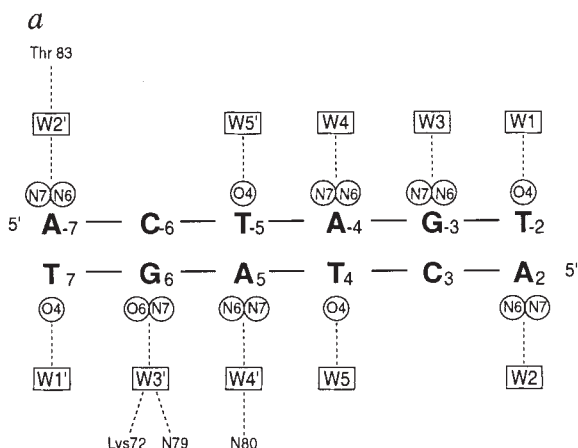
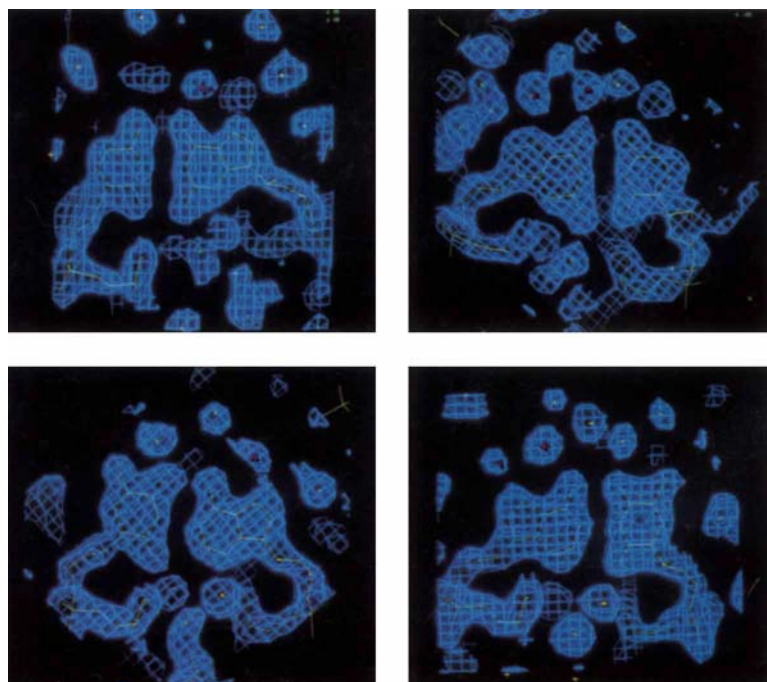
plexed and free target sequences in the *trp* system shows that the three hydration sites used to mediate protein contacts to the three critical bases of the operator's half-site sequence are already fully occupied in the free DNA. These water molecules can thus be regarded as non-covalent extensions of the DNA bases which may be used as stereospecific recognition elements of the DNA target sequence.

Our so-called 'free' DNA is not actually free, because its structure may be influenced by crystal packing forces^{3,12,13}; also, the bases flanking the 6-bp recognition site are different from those in the operator. We cannot estimate the effects of these factors, but the crystal structures of the same 6-bp fragment in the bound and free states, each in different and non-symmetrical environ-

FIG. 3 a, Schematic representation of the 10 conserved hydration sites in the major groove of the *trp* operator half-site in the free and bound states. Water molecules occupying these sites are at hydrogen-bond distances from the donor or acceptor atoms of the DNA bases. Contacts shorter than 3.3 Å can be considered as potential hydrogen bonds²². The water molecules are denoted as W1–W5 for the first 3-bp region and W1'–W5' for the symmetry-related region. Three water molecules in the complex (W2', W3' and W4') mediate four contacts to the protein (Thr 83, Lys 72, N79 and N80). b, Stereo views of the major-groove hydration of the central ACTAGT region in the free decamer (top) and the same segment in one of the four half-sites of the complex (bottom). Only first and second hydration-shell molecules are represented. The water molecules occupying the 10 conserved hydration sites are shown in red; the three molecules mediating contacts to the protein in the complex (W2', W3', W4') and their equivalents in the free decamer are the larger spheres. The two additional water molecules in the doubly hydrated bases A₅ and G₆ of the free fragment (denoted as W3'' and W4'') are also shown in red. All other water molecules are shown in blue. The conserved hydration pattern of the region distant from the protein (W1–W5) is very similar to the equivalent region in the free target. c, 2F_o–F_c electron density maps, calculated at the 4 central base pairs of the 10-mer, showing the various hydration patterns. The lowest contour level is 1.0σ. Top left to right: singly hydrated and doubly hydrated guanines (G₋₃ and G₆). Bottom left to right: singly hydrated and doubly hydrated adenines (A₋₄ and A₅). DNA is in yellow and electron density in blue. Crosses indicate water molecules, those in red being those that form hydrogen bonds to the guanine and adenine bases in the major groove.



c



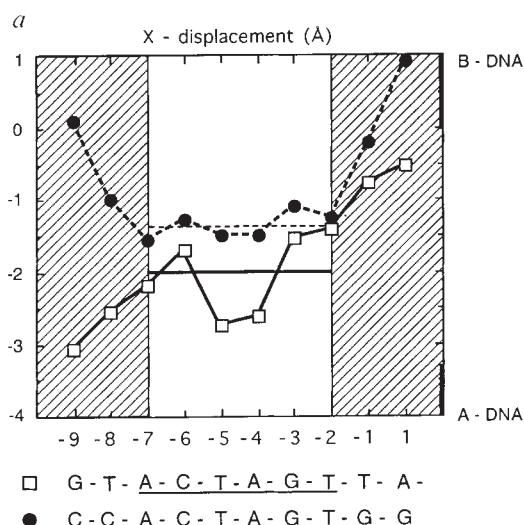
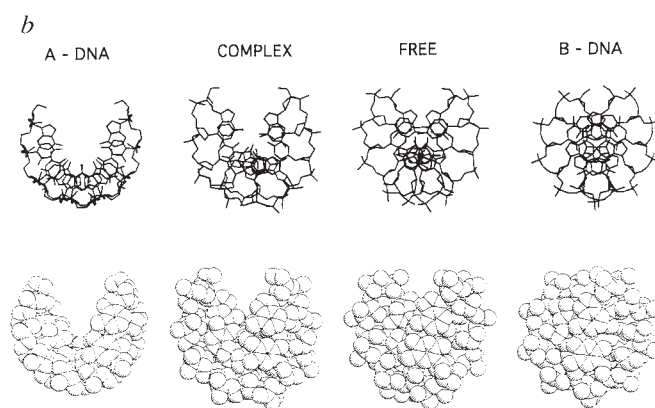


FIG. 2 a, X-displacement plots of the free decamer (circles) and the complexed operator (squares). X-displacement is the perpendicular distance from the C6-C8 vector of a base pair to the helix axis¹⁴. The numbering scheme for the DNA operator half-site is the one used previously and assigns zero to the dyad axis of the *trp* operator¹. The top strand (from 5' to 3' direction and shown at the bottom of the figure) goes from -9 to 1 and that of the complementary strand (not shown) goes from -1 to 9. The same numbering scheme is used for the decamer. Calculations were done with NEWHEL93, provided by R. E. Dickerson. In the free decamer, the calculation of the overall helix axis was based on all 10 bp. In the complexed operator, the local helix based on 10 bp was derived for each of the four crystallographically independent half-sites and the displacement value for each base pair is an average of four independent values (s.d., 0.06–0.26 Å). The regions outside the central 6-bp segment are shaded. Horizontal lines show the corresponding mean values for the ACTAGT regions in the complexed target (continuous line at -2.0 Å) and the free target (broken line at -1.4 Å). The displacement patterns of the complexed and free ACTAGT regions are highly correlated (correlation coefficient, 0.9); that is, the relative displacement of each of the target's 6 bp with respect to its neighbours in the complexed operator is similar to that in the decamer. Whereas the free decamer has a nearly symmetrical pattern, the complex does not, particularly at the terminal base pairs, where

asymmetry is imposed by the flanking base pairs and the interacting protein. b, Four conformations of the double-helical ACTAGT fragment viewed down the helix axis. Both stick and space-filling drawings are shown. The structures of A-DNA and B-DNA were derived from fibre-based coordinates¹⁶ and are referred to as standard forms. The free structure is the central 6-bp segment in the crystal structure of the decamer. The complexed structure is represented by the corresponding segment in one of the four nearly identical half-sites in the crystal structure of the complex. The cavity down the helix axis is indicative of a deep major groove. The depth of the major groove in the complexed fragment is half way between that in A-DNA and B-DNA, and has been found in the crystal structure of a complex between DNase I and a DNA octameric oligonucleotide²¹. The decanucleotide conformation is intermediate between that of the complexed fragment and standard B-DNA. The projected gap between the two 5' ends of the two strands is indicative of the degree of helical unwinding: 11 bp per turn for standard A-DNA, 10.9 for the complexed half-site operator, 10.6 for the same segment in the decamer, and 10 for standard B-DNA. The only other cases in which the helices are underwound (10.5 and 10.6 base pairs per turn) are the two DNA decamers CCACTTGG and CGATCGATCG, which crystallize with packing similar to the present decamer^{2,3}. But these two decamers and the *trp* 10-mer differ in their local conformations and other overall features: for example, the central 6-bp regions of the three helices (AACITT, ATCGAT and ACTAGT) show displacement patterns that are distinctly different. The average X-displacement values are -0.5, -0.6 Å and -1.4 Å, respectively, indicating that the deep major groove in the unbound *trp* hexamer is an intrinsic feature of the particular sequence.



asymmetry is imposed by the flanking base pairs and the interacting protein. b, Four conformations of the double-helical ACTAGT fragment viewed down the helix axis. Both stick and space-filling drawings are shown. The structures of A-DNA and B-DNA were derived from fibre-based coordinates¹⁶ and are referred to as standard forms. The free structure is the central 6-bp segment in the crystal structure of the decamer. The complexed structure is represented by the corresponding segment in one of the four nearly identical half-sites in the crystal structure of the complex. The cavity down the helix axis is indicative of a deep major groove. The depth of the major groove in the complexed fragment is half way between that in A-DNA and B-DNA, and has been found in the crystal structure of a complex between DNase I and a DNA octameric oligonucleotide²¹. The decanucleotide conformation is intermediate between that of the complexed fragment and standard B-DNA. The projected gap between the two 5' ends of the two strands is indicative of the degree of helical unwinding: 11 bp per turn for standard A-DNA, 10.9 for the complexed half-site operator, 10.6 for the same segment in the decamer, and 10 for standard B-DNA. The only other cases in which the helices are underwound (10.5 and 10.6 base pairs per turn) are the two DNA decamers CCACTTGG and CGATCGATCG, which crystallize with packing similar to the present decamer^{2,3}. But these two decamers and the *trp* 10-mer differ in their local conformations and other overall features: for example, the central 6-bp regions of the three helices (AACITT, ATCGAT and ACTAGT) show displacement patterns that are distinctly different. The average X-displacement values are -0.5, -0.6 Å and -1.4 Å, respectively, indicating that the deep major groove in the unbound *trp* hexamer is an intrinsic feature of the particular sequence.

21. Weston, S. A., Lahm, A. & Suck, D. *J. molec. Biol.* **226**, 1237–1256 (1992).
22. Schneider, B., Cohen, D. & Berman, H. M. *Biopolymers* **32**, 725–750 (1992).

ACKNOWLEDGEMENTS. This work was supported by the United States–Israel Binational Science Foundation, the Israel Science Foundation administered by the Israel Academy of Sciences and Humanities, the Helen and Milton Kimmelman Center for Macromolecular Assembly, and the NIH. We thank A. Klug and M. Levitt for discussion. Z. S. holds the Helena Rubinstein Professorial Chair of Structural Biology. D. R. holds the William and Lee Abramowitz Professorial Chair of Macromolecular Biophysics.

Received 6 May 1993; accepted 8 February 1994.

- Otwinowski, Z. et al. *Nature* **335**, 321–329 (1988).
- Lipmanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J. & Dickerson, R. E. *Biochemistry* **32**, 1373–1389 (1993).
- Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J. & Dickerson, R. E. *J. molec. Biol.* **231**, 768–784 (1993).
- Privé, G. G., Yanagi, K. & Dickerson, R. E. *J. molec. Biol.* **217**, 177–199 (1991).
- Heinemann, U. & Alings, K. *J. molec. Biol.* **210**, 369–381 (1989).
- Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. *J. biol. Chem.* **266**, 8861–8883 (1991).
- Quintana, J., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. *J. molec. Biol.* **225**, 379–395 (1991).
- Yuan, H., Quintana, J. & Dickerson, R. E. *Biochemistry*, **31**, 8009–8021 (1992).
- Shakked, Z. & Rabinovich, D. *Prog. Biophys. molec. Biol.* **47**, 159–195 (1986).
- Travers, A. A. & Klug, A. in *DNA Topology and its Biological Effects* (eds Cozzarelli, N. R. & Wang, J. C.) 57–106 (Cold Spring Harbor Laboratory Press, New York, 1990).
- Sigler, P. B. in *Transcriptional Regulation* (eds McKnight, S. L. & Yamamoto, K.) Vol. 1, 475–499 (Cold Spring Harbor Laboratory Press, New York, 1992).
- Shakked, Z., Guerin-Guzikovich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. *Nature* **342**, 456–459 (1989).
- Shakked, Z. *Curr. Opin. struct. Biol.* **1**, 446–452 (1991).
- Dickerson, R. E. et al. *EMBO J.* **8**, 1–4 (1989).
- Rabinovich, D. & Shakked, Z. *Acta crystallogr.* **A40**, 195–200 (1984).
- Chandrasekaran, R. & Arnott, S. in *Landolt-Bornstein New Series, Group VII* (ed. Saenger, W.) Vol. 1b, 31–70 (Springer, Berlin 1989).
- Brünger, A. T. *X-PLOR Version 3.0: A system for Crystallography and NMR* (Yale Univ., 1992).
- Hendrickson, W. A. & Konert, J. H. *Biomolecular Structure, Conformation, Function and Evolution* (ed. Srinivasan, R.) Vol. 1, 43–47 (Pergamon, Oxford, 1981).
- Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119–145 (1985).
- Timis, Y., Westhof, E., Fuchs, P. P. & Moras, D. *Nature* **361**, 459–462 (1989).

ERRATUM

High-affinity IgE receptor on eosinophils is involved in defence against parasites

Abdellilah Soussi Gounni, Bouchaib Lamkhoud, Kenichi Ochial, Yoichi Tanaka, Emmanuel Delaporte, André Capron, Jean-Pierre Kinet & Monique Capron

Nature **367**, 183–186 (1994).

IN Fig. 4a and b of this letter, the ordinate axes were both labelled incorrectly: they should read “EPO index” and “ECP index”, respectively.

A model for high-throughput automated DNA sequencing and analysis core facilities

Mark D. Adams, Anthony R. Kerlavage, Jenny M. Kelley, Jeannine D. Gocayne, Chris Fields, Claire M. Fraser and J. Craig Venter

Large-scale DNA sequencing and analysis is providing important molecular information on complex biological systems.

RAPID gene discovery through automated DNA sequencing is accelerating the pace of biological and medical research¹⁻⁹. Automated sequencing instruments were first described in 1986 (ref. 10) and became commercially available in the late 1980s. These instruments combine modifications of standard polyacrylamide gel electrophoresis used in manual DNA sequencing with real-time detection of fluorescence-labelled DNA fragments. Their advantage comes both from automated data collection and determination of nucleotide sequence (base-calling) and from the increased number of samples that can be processed. In February 1987, we helped implement this technology¹¹, a critical step required for the creation of a high-throughput DNA sequencing laboratory, by becoming a test site for the first automated sequencer from Applied Biosystems, Inc. (ABI, now the Applied Biosystems Division of Perkin-Elmer, Foster City, California).

Between 1987 and 1991, while at the US National Institutes of Health (NIH), we demonstrated the feasibility of increasing throughput in a parallel fashion by expansion from one to six automated sequencers and addition of two ABI Catalyst robots¹². This development enabled the sequencing of the entire smallpox virus genome⁸, six human cosmids¹³⁻¹⁴, and 11,000 human and *Caenorhabditis elegans* expressed sequence tags (ESTs)^{1-4,15} over the three-year period from 1990 to 1992. Our NIH sequencing laboratory provided a model upon which many other automated sequencing facilities have been based.

The Institute for Genomic Research (TIGR) was founded in July 1992, as a nonprofit research institute, with enlarged sequencing and computational core facilities for a sequence-based characterization of gene expression and molecular phylogeny. The design of the DNA sequencing laboratory was based around thirty ABI 373A automated DNA sequencers

and seventeen ABI 800 Catalyst machines (see figure).

Guiding principles

Two principles guided the design of the TIGR core facility and continue to guide its operation: flexibility to adopt improvements at each stage and the integration of data analysis with all phases of the sequencing process. The facility can accommodate a wide variety of projects, ranging from complementary DNA to genomic sequencing. From our first effective utilization of automated DNA sequencing¹¹ to the present, there has been a continuous evolution in

strive to maintain a modular system, into which new protocols, instrumentation, or approaches can be substituted with minimal disruption to the overall organization.

Template preparation, sequencing reactions, and gel electrophoresis are routinely performed on 960 templates per day by a staff of 27 skilled technicians. Following gel electrophoresis and automated data collection, each of the electropherograms is inspected visually, unreliable sequence is deleted from the files, and the sequences and related information are loaded into a custom relational database, ESTDB¹⁶. As outlined in the table, the technicians are grouped into

teams based around parts of the sequencing process or instrument groups. In a laboratory where the 'production' aspects are easily quantified, the balance between specialization and diversity is particularly important in order to recruit and retain talented technicians who have the ability and flexibility to solve technical problems and to participate in reaching the scientific goals of the sequencing projects.

Integrating informatics

Maintaining a steady state data flow when dealing with nearly a thousand samples per day requires a consistency in methods and laboratory practice that is unusual in basic research laboratories. Data analysis and management become integral parts of the sequencing process, and bottlenecks at any stage have the potential to disrupt or delay the entire process. Delays in sequence data analysis, which provides feedback on library and sequence quality, can affect decisions taken later on, such as choice of library or screening procedures that have a significant impact on the overall productivity of the project.

ESTDB, implemented using the commercial system Sybase (Sybase, Inc., Emeryville, California), maintains all data from the sequencing process, from library sources to sequence analysis results. At each step, information, such as reaction conditions, operator, and lot number of reagents is



The DNA sequencing core facility at TIGR.

DNA template preparation methods, sequencing chemistry, instrumentation, data tracking and software for sequence analysis. For example, four different template preparation methods have been used over the past year, ranging from the polymerase chain reaction to plasmids, each of which has special requirements for downstream sequencing reactions. Expansion of the amount of sequence data obtained is possible with existing equipment by increasing the number of lanes run per gel, decreasing electrophoresis time and extending sequencing read length with Stretch Liners. (The Stretch Liner is a modification of the ABI 373A sequencer that increases the electrophoresis distance prior to fluorescence detection, thereby producing longer sequence read lengths.) We

TIGR photo file

Current staff of the TIGR sequencing facility

Group	No. of teams	No. of people	Job description
Gel	1	4	Wash and tape all sequencing gel plates, pour all gels for 30 sequencers daily
Template	1	4	Prep up to 1,200 templates per day, including clone picking, quantitation, clone replication
Sequencer-A	4	3-4	Run three Catalysts and six sequencers, data editing, quality control
Sequencer-B	1	3	Sequencer-A plus fast turnaround, training and special projects

collected and stored in ESTDB. Internal integrity checks are performed automatically to assure data consistency. Because ESTDB is the repository of all sample tracking and analysis results, and because most of our data tracking and analysis tools interface with it, ESTDB is implemented on a high-performance Sun SPARCcenter2000 computer (Sun Microsystems, Inc., Mountain View, California).

A high-capacity ethernet network was installed to link over 100 computers at TIGR and to handle the high volume of data generated. Sequence electropherograms alone comprise over 150 megabytes of data, and therefore network traffic, daily. Electropherograms are stored on a central Macintosh Workgroup Server 95 (Apple Computer, Inc., Cupertino, California) with six 1-gigabyte hard disks, where they are edited and then archived to 8-mm tape. Multiple tape copies are maintained in different locations to ensure the safety and security of the data.

Effective analysis is essential to extract as much information as possible from the sequence data. Thus, most of the analysis process for ESTs has been highly automated. We have developed an application (S.T.P., A. R. Kerlavage *et al.*, unpublished data) which extracts sequences from ESTDB based upon predefined queries, and provides the user with a selection of analysis routines to apply to the sequences. The routine processing steps for new sequences are as follows: (1) sequences are checked for vector at both ends; (2) Alu repeats and other common repetitive elements are 'masked' from the sequence so that portion is not used in further processing; (3) ESTs are searched against a 'prescreen' file of curated human sequences, if one is matched exactly, a name is automatically assigned to the EST; (4) nucleotide and protein searches are performed against non-redundant databases. We run BLASTN¹⁷ (NCBI, Bethesda, Maryland) locally on a Sun SPARC10 (Sun Microsystems) for nucleotide searches and run BLAZE¹⁸ (Intelligenetics, Inc., Mountain View, California) on TIGR's 4000 processor MasPar 2204 (MasPar, Inc., Sunnyvale, California) for protein searches. In addition, all libraries are checked using a hexamer-

based quality control algorithm¹⁹, which provides a measure of likelihood that a given sequence originated from a particular species. This method is particularly effective in detecting library contamination. Libraries that demonstrate such contamination, for example, by *Escherichia coli*, or yeast, are discarded and the sequences from them are not used for further study.

Following this automated series of analyses, scientists examine the nucleotide and protein alignments using a program (BYOB, A. R. K. *et al.*, unpublished data) which allows the user to view the alignments, select a match to be saved to the database, and assign an identification to the EST. These are often judgement calls that require a background and experience in evaluating sequence similarity.

With all of the information stored in ESTDB, correlations can be constructed and tested by querying the database. It is important to monitor the quality at each step, as sacrifices in quality will be amplified throughout the process. Furthermore, the most valuable information from EST projects is not necessarily a list of genes matched but the answers to more complex questions such as: 'How do library A and library B from the same tissue differ in their representation of abundant ESTs?' or 'In which library is gene X most abundant?'

Discussion

The key to successful large-scale sequencing, regardless of the particular application, is balance. If data analysis is held until the end, systematic problems with the data will not be discovered until considerable time and money have been spent; if a fivefold improvement in sequencing throughput cannot be supported by a fivefold increase in template preparation and data analysis, then the advantages cannot be utilized. There is no such thing as a static production line for sequencing. Every step must be continually monitored and analysed, both for consistency of data and opportunities for improvements, by knowledgeable, trained scientists. □

Mark D. Adams, Anthony R. Kerlavage, Jenny M. Kelley, Jeannine D. Gocayne, Chris Fields,

Claire M. Fraser and J. Craig Venter are at The Institute for Genomic Research, 932 Clopper Road, Gaithersburg, Maryland 20878, USA. We thank the TIGR Sequencing and Computational Core, and scientists in the Departments of Genome Informatics, Cellular and Molecular Biology, and Gene Discovery and Evolution. For more information, fill in reader service number 100.

- Adams, M. D. *et al.* *Science* **252**, 1651-1656 (1991).
- Adams, M. D. *et al.* *Nature*, **355**, 632-634 (1992).
- Adams, M. D. *et al.* *Nature Genet.* **4**, 256-267 (1993).
- Adams, M. D. *et al.* *Nature Genet.* **4**, 373-380 (1993).
- Khan, A. S. *et al.* *Nature Genet.* **2**, 180-185 (1992).
- Durkin, A. S. *et al.* *Cytogenet. Cell Genet.* **65**, 86-91 (1994).
- Polymeropoulos, M. H. *et al.* *Nature Genet.* **4**, 381-386 (1993).
- Massung, R. F. *et al.* *Nature* **366**, 748-751 (1993).
- Papadopoulos, N. *et al.* *Science*, **263**, 1625-1629 (1994).
- Smith, L. M. *et al.* *Nature* **321**, 674-678 (1986).
- Gocayne, J. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 8296-8300 (1987).
- Cathcart, R. *Nature* **347**, 310 (1990).
- McCombie, W. R. *et al.* *Nature Genet.* **1**, 348-353 (1992).
- Martin-Gallardo, A. *et al.* *Nature Genet.* **1**, 34-39 (1992).
- McCombie, W. R. *et al.* *Nature Genet.* **1**, 124-131 (1992).
- Kerlavage, A. R. *et al.* *Proc. 26th Hawaii int. Conf. System Sciences*, 585-594 (IEEE Computer Society Press, Los Alamitos, California, 1993).
- Altschul, S. F. *et al.* *J. molec. Biol.* **215**, 403-410 (1990).
- Brutlag, D. L. *et al.* *Comp. Chem.* **17**, 203-207 (1993).
- White, O. *et al.* *Nucleic Acids Res.* **21**, 3829-3838 (1993).

ADVERTISEMENTS

DNA SEQUENCING

Using an ABI 373 Automated Sequencing System
£60/sample

- PCR products
- M13 inserts
- plasmid inserts
- 300-400 bases of sequence
- single stranded DNA
- double stranded DNA
- solid phase sequencing

Please contact us for more information and a preparation protocol

Molecular Medicine Unit

King's College School of Medicine & Dentistry

Tel: 071-326 3126 Fax: 071-733 3877

Reader Service No.10

OLIGONUCLEOTIDES

Synthesised, purified and shipped in 3 days

£2.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry

Tel: 071-326 3126 Fax: 071-733 3877

Reader Service No.11

Recent releases

New product developments to watch for include a whole chromosome painting system, a second-generation robotics system, a pocket-sized torch for use in the darkroom and two new gel documentation systems.

AMERSHAM's range of **cDNA rapid cloning modules** now includes the two phage vectors, lambda-MOSSlox and MOSElox (*Reader Service No. 101*). Intended for cDNA library construction of high-efficiency and low non-recombinant background, both vectors accept inserts of up to 8 kilobases in length and feature an advanced automatic subcloning system for the conversion of phage recombinants to plasmids *in vivo*. Complete plasmids are built into the phage and are flanked by two 34 base pairs



Phage cloning vectors.

loxP sites derived from bacteriophage P1. Plasmid subclones are generated by infection of hosts expressing the P1 cre recombinase, which recognizes the loxP sites and forms the plasmid by site-specific recombination. When plated in the presence of ampicillin, colonies appear which are the result of plasmid excision. Amersham says that inserts are faithfully and efficiently subcloned without the generation of deletions or rearrangements: this is in contrast to systems employing single-stranded replication intermediates such as those based on M13-type helper phage infection.

Imagenetics has developed a new **whole chromosome painting system** for use in fluorescence *in situ* hybridizations (*Reader Service No. 102*). The company has also added two new probes, SpectrumGreen WCP 18 and SpectrumGreen WCP X, to its line of chromosome probes. The WCP system is designed to enable researchers to identify human chromosome aberrations in metaphase spreads and to verify the chromosomal origin of marker chromosomes. Each probe is covalently linked to the company's

SpectrumOrange or SpectrumGreen fluorophores, enabling DNA targets to be simultaneously hybridized and labelled. Because most of the probes are available with either fluorophore, samples may be screened with multiple probes in one assay. All WCP probes contain unique sequences homologous to DNA along the length of the target chromosomes, fluorescently "painting" the whole specific chromosomes of interest. The directly-labelled WCP system is said to eliminate many of the time-consuming steps of indirectly-labelled methods, reducing the likelihood of nontarget fluorescence and producing results that have sharp signals. Stated applications include the detection of chromosome translocations, identification of marker chromosomes, identification of chromosomal changes due to chemical mutagenesis, scoring of chromosomal aberrations in radiation dosimetry studies, identification of human chromosomes in somatic cell hybrids and mapping human genes.

Biomek 2000 is Beckman Instruments' second-generation **birobotics system for laboratory automation** (*Reader Service No. 103*). The system can be used to automate



Beckman — making life easier.

assays and sample preparation procedures including such tasks as pipetting, diluting, plate washing, high-density replicating and the quantification of photometric assays. It should find application in a wide range of scientific disciplines including biochemistry, clinical research, drug discovery, immunology, molecular biology and molecular diagnostics. The workstation features a worksurface that has up to 12 selectable positions and so can be configured to suit the user's needs and choices of labware. The graphic displays of the Windows (Microsoft)-based BioWorks software are designed to guide quickly the user through the selection of labware and the configuration of assays, while providing access to optional accessories and other software programs. Methods are created and editing is simplified by point-and-click operation using icons. The workstation features interchangeable liquid-handling tools that allow the user to

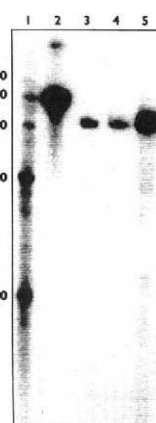
transfer volumes from 1 µl to many millilitres. Pipetting tools, including those with sonic liquid-level sensing, use a selection of disposable pipette tips that range from standard to microvolume, aerosol-resistant to nonpyrogenic.

When coupled with the biotin/streptavidin chemiluminescent western blotting kit, the **nonradioactive *in vitro* translation kit** from Boehringer Mannheim is designed to provide sensitivity, speed and safety during protein synthesis (*Reader Service No. 104*). The mRNA is translated into a biotinylated protein using the reticulocyte lysate and charged lysine-tRNA supplied. Following the translation reaction, separation by polyacrylamide gel electrophoresis, and western blotting, the protein can be detected in under 2 hours and with a sensitivity that is said to be comparable to radiolabelled proteins. Using these kits, Boehringer Mannheim says that the procedure of translation, separation and detection can be accomplished within 6 hours; this compared to the 8–14 hours needed for radioactive procedures. In addition, the biotin-labelled proteins are not degraded by autoradiolysis, eliminating the need for repeat labelling and the additional exposure to and expense of radioactivity.

Assays

The Direct Protect — Lysate RPA — **ribonuclease protection assay kit** from Ambion is designed for the direct detection, quantitation and characterization of RNA in crude cell lysates from cultured cells or solid tis-

Mouse kidney and liver tissue, and cultured (BW5147) cells were lysed and probed directly in the lysate with a radioactive antisense RNA probe and processed with the kit. A 316-base antisense GAPDH transcript was used as a probe in all samples. Lane 1, RNA markers; lane 2, minus RNase control; lane 3, kidney; lane 4, liver; lane 5, tissue culture cells.



issues without the need for prior RNA isolation (*Reader Service No. 105*). With this assay an RNA probe is hybridized directly with cellular RNA in a crude guanidine thiocyanate cell lysate. (Guanidine thiocyanate is a chaotropic salt that rapidly solubilizes cells and tissues while inactivat-

ing cellular ribonucleases. Additionally, it is said to be an efficient environment for nucleic acid hybridization.) After hybridization, excess probe and sample RNA are removed by treating with ribonucleases and the hybrids precipitated. Ambion says that the entire procedure can be carried out in a single tube. The protected probe hybrids are then separated on a denaturing polyacrylamide gel and visualized by autoradiography. This kit should prove useful in situations that require a large number of samples to be analysed, or where a yes/no answer is appropriate. The stated detection limit for the assay is about 30 femtograms of target RNA.

Trevigen, expanding its interest in tissue-specific gene expression, has developed a two-part system that consists of the mRNA GENERATOR and mRNA protection assay kits (*Reader Service No. 106*). With this new **mRNA protection assay system**, the company says that researchers can detect and quantitate mRNA directly from whole cell lysates, eliminating the need to purify RNA. The system works with either whole cells or tissues and allows a researcher to make a radiolabelled antisense probe to the mRNA of interest, hybridize to a target in a whole cell lysate, remove the unhybridized probe and message and present only the hybrid for subsequent detection and quantitation. Said to be capable of detecting less than 0.1 pg of mRNA per reaction, the new system provides enough reagents to carry out 100 protection assay reactions.

■ Primers/probes and peripherals

For the **automatic cleavage, deprotection and concentration of synthesized DNA**, Savant has developed the OP120 OligoPrep benchtop system which is available from Life Sciences International (*Reader Service No. 107*). Weighing in at just 49 kg and measuring 42.5 × 35.6 × 65.5 cm, the OligoPrep combines a rotor and vial/cap, accommodating up to 12 individual, iso-



Savant's programmable concentrator.

lated sample assemblies. These assemblies remain closed during the incubation at 55 °C, and automatically open by centrifugal force during the evaporation stage. Both incubation and concentration times can be programmed using the instrument's front panels; the instrument can also be preprogrammed and left unattended for overnight processing. A corrosion-resistant stainless steel vapour shaft and an oil-free vacuum pump are designed to reduce maintenance and extend the life of the instrument. An additional post-trap is said to provide odour-free neutralization of ammonia and disposal of solvents.

Spot Checked Oligos is the new purity check kit from Severn Biotech that is designed to provide a quick **spot-check of the purity of an oligonucleotide** prior to any further purification or post-synthesis handling procedures (*Reader Service No. 108*). The kit enables the user to separate sequences up to 60 bases in length. The company says that as most oligos are within the 25–45mer band length, the kit is designed to provide optimal resolution with sequences of this length. Oligos can be loaded using a variety of different solvent solutions, which enables them to be separated in the deprotection ammonia solution directly after synthesis. A minimum of 5 µg DNA is required to perform the test. Results can be obtained within 90 min.

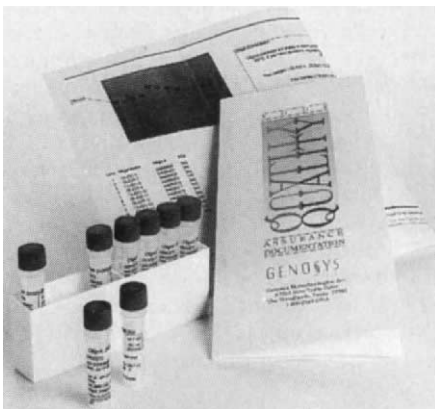
Advanced Biotechnologies announces the availability of 11 **primer kits for random amplified polymorphic DNA analysis** (*Reader Service No. 109*). The kits contain



Advanced Biotechnologies' primer kits.

twenty 10-base oligonucleotide primers for use in genetic mapping and DNA fingerprinting. In these studies, genomic DNA from various individuals will often produce different amplification fragment patterns when analysed by electrophoresis. The differences in these patterns represent polymorphisms that can be used as genetic markers. The bases in each primer are selected randomly with the requirement that their (G + C) ratio is 60 to 70 per cent, and that they do not have self-complementary ends. The primers are desalted by NAP-10 columns and contain approximately 15 µg — said to be sufficient for about 1,000 amplification reactions.

Genosys offers kits of small primers useful for **DNA fingerprinting and identification of genetic markers** (*Reader Service No. 110*). The company says that the arbitrary 10mer sequences have been selected on the basis of their percentage GC content and are without palindromic sequences or complementary ends. Kits with 50 to 80 per cent GC content primers are available, as well as ones containing primers that incorporate common restriction sites, enabling



10mer kits for DNA fingerprinting.

amplified products to be subsequently cloned. The company has also introduced a range of **labelled primers for DNA sequencing**. The common sequencing primers M13 (-21) Universal, M13 Reverse, T3, T7 and SP6 primers are available conjugated to biotin or fluorescein. The fluorescein primers can also be used in Pharmacia's automated DNA sequencer. Sequencing primers are also available labelled with 4 ABI dyes (ROX, JOE, FAM and TAMRA) and supplied as a kit for use with Applied Biosystems' automated DNA sequencer. Enough primer for over 100 double-stranded or single-stranded sequencing reactions is provided.

Bios has introduced a set of thirteen oligonucleotides that may be used as **probes to screen cloned DNA** for the presence and type of the common occurring short tandem repeats (STR) or microsatellites (*Reader Service No. 111*). Microsatellites consist of sequence motifs from 1 to 6 base pairs that may occur as tandem repeats. They occur randomly and frequently in all eukaryotic DNAs, except yeast. In humans, 76 per cent of the repeated motifs are A, AC, AAAN, AAN or AG (listed in decreasing order of abundance). The Bios system includes these sequence motifs. The oligonucleotides are said to have similar thermal melting properties, which facilitates screening hybridization with a single set of conditions (calculated annealing temperature 72 °C). The probes can be labelled by radioactive or nonradioactive methods. The company also offers nine repetitive element primers and nine vector primers. The former consists of six Alu primers, one long interspersed repeat element [LINE (L1)] primer, and two telomere primers. The vector primers con-

PRODUCT REVIEW

sist of M13 forward and reverse sequencing primers, lambda gt 10 and gt 11 forward and reverse primers, T7 promoter primer, SP6 promoter primer and the T3 promoter primer.

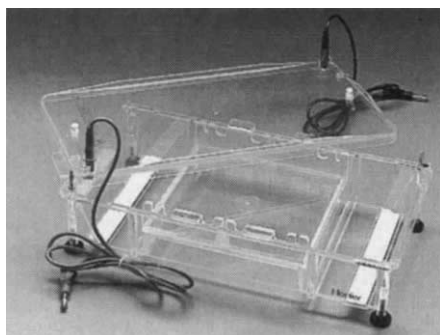
■ RNase Inhibitors

Eliminating endogenous RNases from biological materials can be a headache but Amresco thinks it may have the answer. The company is now offering a **human placental RNase inhibitor**, supplied as a highly purified, nuclease-free preparation, which is said to be effective against RNase A and other eukaryotic RNase A-like proteins (*Reader Service No. 112*). Because the preparation does not inhibit RNA polymerases, DNA polymerases, RNase H or RNase T1 it is therefore suitable for use in rtPCR reactions, *in vitro* transcription or in the construction of cDNA libraries. Supplied at a concentration of 40,000 units per millilitre in a buffer containing 20 mM HEPES-KOH, 50 mM KCl, 5 mM DTT and 50 per cent glycerol, one unit is said to inhibit 50 per cent of the activity of 5 ng RNase A.

Eliminate unwanted RNase and DNA contaminants from laboratory plastic and glassware with the new RNase Away reagent from Molecular Bio-Products (*Reader Service No. 113*). This **ribonuclease decontaminant** is said to be nonhazardous and is supplied in a ready-to-use form. Simply apply RNase Away to the surface to be decontaminated and rinse, says the company.

■ Electrophoresis systems

Hoefer Scientific's new HE 99X Max **horizontal agarose unit** is designed to offer maximum resolution of DNA fragments on standard gels, those that are 15-cm wide and 10-, 15- or 20-cm long (*Reader Service No. 114*). A variety of gel lengths and comb



Hoefer's Scientific: taking it to the Max.

sizes are offered. Other design features include leak-proof casting trays with foam-sealed ends, self-locating centring guides on running trays, banana plugs that are corrosion-resistant, an electrode carrier for simple electrode replacement, a rugged moulded lid with safety guards to protect against electrical hazards, strain-relieved high-voltage cords to prevent lead breakage, red and black markers on both running trays and buffer chambers for quick orientation.

478

Pharmacia has a new range of EPS **electrophoresis power supplies** (*Reader Service No. 115*). Three models are available: each is stackable with two sets of outputs. The EPS 600 (600 V, 400 mA, 100 W) can store three programmed methods and is intended for standard protein and DNA applications. For DNA sequencing applications and isoelectric focusing, the company recommends its EPS 3500 (3,500 V, 150 mA,



Pharmacia's top-of-the-line model.

100 W) power pack, which can store up to three programmed methods. The deluxe model, the EPS 3500XL (3,500 V, 150 mA, 100W), can store nine programmed methods, each of up to nine phases. The EPS 3500XL can be fine-tuned for such demanding techniques as two-dimensional electrophoresis and multi-step isoelectric focusing.

The new CE2000 automated **capillary electrophoresis system**, manufactured by Advanced Molecular Systems, is now available through GSG Interion (*Reader Service No. 116*). The system utilizes sample and capillary thermostating. It also offers constant current or constant voltage control with multilinear ramping, in addition to a fully programmable 30 kV power supply for high-resolution separations. An autosampler allows the researcher automatically to run up to 96 samples. Samples can be injected by either electrokinetic or controlled vacuum methods. Up to nine different methods can be used with up to four different buffers. Two detection modes are available: a multichannel ultraviolet photodiode array detection system with up to four outputs selected from five fixed wavelengths between 197 nm and 280 nm, and a multichannel fluorescence photodiode array detection system with up to four outputs selected from 12 fixed emission wavelengths between 457 nm and 621 nm.

■ Handy gadgets

Amersham's new handy-sized, **red light emitting pocket torch**, Hypertorch, is designed for use in dark rooms with autoradiography films and emulsions (*Reader Service No. 117*). The device provides the minimum safelight exposure when handling very sensitive emulsions. It also helps in the location of equipment in areas of the dark-room poorly illuminated by the safelight.

Eliminate stubborn air bubbles from a gel in one easy step, says AT Biochem, with

its new BubbleBuster **electrophoresis tool** (*Reader Service No. 118*). These new gadgets are specially designed for sequencing plates with 0.4-mm spacers and are supplied five to a pack.

■ Documentation

DOC-IT is the new **gel documentation system** from Ultra Violet Products that is based on the company's GDS 5000 system and which is being pitched as an entry-level documentation device (*Reader Service No. 119*). The system uses a high-resolution CCD camera to capture images and to transfer them to the processing unit, in addition to a black and white monitor with simple-to-use camera controls that allow operation of the integration functions of the camera. A 3UV transilluminator that generates 365, 302 and 254 nm is optional. The filter area is 20 × 20 cm, which should suit most gel sizes. It is possible to print from DOC-IT by connecting a thermal printer such as the Sony UP 860 or any similar device. Printer options for dot matrix or laser printing will be available shortly.

MiniVisionary **gel documentation system** from Fotodyne, which is said to cost under 10-cents-a-photograph to operate (*Reader Service No. 120*). The system is



MiniVisionary: for the cost conscious.

compatible with most Fotodyne and Polaroid hoods. Photographs are digitally produced using CCD technology and are output to a video printer. The system is intended for applications in research and education.

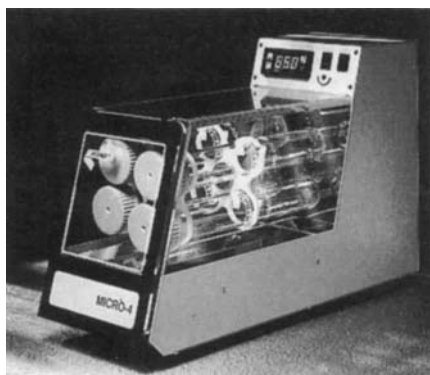
■ Assorted hardware

GeneE is a self-contained **thermal cycler** that uses a recirculated coolant to provide maximum cooling efficiency and which has a stated cooling rate of 2.7 °C per second (*Reader Service No. 121*). The instrument operates from 10 °C above ambient to 99 °C, and can store up to 99 separate programmes. Selectable ramp rates can be varied from 1 °C to 40 °C per minute and can be displayed in either °C or °F. Additional features include a pause button for stopping the pro-

NATURE · VOL 368 · 31 MARCH 1994

gramme or for manually denaturing samples, and an audible bleeper.

Hybaid is offering the Micro 4 **personal hybridization system** (Reader Service No. 122). This compact four-bottle unit features



Take four — see Hybaid.

digital temperature control, 10-mm acrylic shielding and a rotisserie design. In addition to the hybridization system, Hybaid has now eliminated the need for oil overlays with its OmniGene multi-format **thermal cycler** now that the device is available with a Hot Lid. With OmniGene, researchers can run up to three independent blocks for 0.5-ml tubes, 96-well plates or slides. The personal data card allows users to save run protocols.

Reagents

Uracil-DNA glycosylase hydrolyses the uracil base from both single- and double-stranded DNA containing deoxyuridine in place of thymidine. After uracil residues have been removed, the deoxyribose sugar backbone can be degraded by exposure to a temperature of 95 °C for 15 min. This treatment also inactivates the enzyme. Uracil-DNA glycosylase does not affect DNA without uracil residues or dUTP, and is inactive on RNA templates. Epicentre Technologies' preparation of ultra-pure **uracil-DNA glycosylase** is said to be free of DNA exo- and endonuclease, RNase and phosphatase, and contaminating apyrimidinic endonuclease activity (Reader Service No. 123).

Tbr polymerase is a new **thermostable enzyme** from NBL Gene Sciences that is said to provide better fidelity than *Taq* polymerase with an error frequency twofold lower (Reader Service No. 124). The company says that *Tbr* polymerase has a half-life of 2.5 h at 96 °C, allows longer primer extensions to be obtained (up to 6,000 base pairs), and has been extensively tested for DNA sequencing.

Catalogues

The new 88-page 1994 catalogue from Invitrogen contains details about the company's 250 **products for molecular biology**, including new products for protein expression and electroporation (Reader Service No. 125). List-

ings include products for the isolation of RNA/DNA, cDNA cloning and analysis, polymerase chain reaction, expression systems and vectors, competent cells, and custom expression and library services.



Invitrogen

Copies of the 1994-95 **catalogue for life science research products** from United States Biochemical are now available and serve as a replacement for both the 1992 molecular biology reagents/protocols catalogue and last year's life science research products catalogue (Reader Service No. 126). The catalogue is divided by product application and has several separate sections for specific product groups such as Ultrapure reagents, restriction enzymes, modifying enzymes and antisera. The catalogue is also available on diskette (Macintosh, DOS or Windows formats).

ICN Biomedicals has introduced a new 24-page **electrophoresis reagent source catalogue** that contains all reagents commonly used for both agarose and polyacrylamide gel electrophoresis techniques (Reader Service No. 127). Products are arranged by category, such as buffers, detergents, stains, agaroses and acrylamides. Also featured are many ready-to-use premixes, including premixed buffers, acrylamide-bis powders and solutions, and saturated phenol:chloroform solutions.

Hoefer Scientific's complete line-up of **electrophoretic equipment** is featured in the company's new 1994 catalogue (Reader Service No. 128). Instruments and applica-



Hoefer Scientific

tions for horizontal submarine electrophoresis, isoelectric focusing, nucleic acid sequencing, preparative electrofocusing, vertical slab-gel units and two-dimensional systems are listed. Related instruments include products for blotting and hybridization, densitometry, filtration, fluorometry, as well as gel dryers, gradient makers and ultraviolet light tables.

More than 2,000 products — over 150 of which are new — are featured in the 1994 **biochemical reagent catalogue** from Boehringer Mannheim (Reader Service No. 129). This new edition contains a number of enhancements: tables for restriction endonucleases, the percentage activity of the company's restriction enzymes in each of the Sure/Cut buffers, a listing of restriction

enzymes and their isoschizomers, and the number of recognition sites on various DNAs. Plasmid maps are provided that indicate the location of the origin of replication, restriction sites, promoter sites and drug-resistance genes, as well as tables that list the position of cleavage sites for known restriction enzymes cutting the plasmid.

A wide selection of ultra-pure **chemicals for biotechnology** are featured in the new 1994-95 Amresco catalogue (Reader Service No. 130). This year, the catalogue features a technical information and application section with numerous protocols. New product listings include a complete line of electrophoresis reagents, nucleic acid purification kits and pre-mixed buffers.



Amresco

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more details on the products, fill in the reader service card bound inside the journal. The polymerase chain reaction (PCR) is covered by patents owned by Hoffman-La Roche, Inc.

ADVERTISEMENTS

Yeast Geneticists

MSM System & MSM Manual

- Ascus Dissection
- Pedigree Analysis
- Cell & Zygote Isolation

SINGER
INSTRUMENTS

"A responsibility to Science"

Tel: (+44) 984 40226 Fax: (+44) 984 41166

Singer Instrument Co. Ltd., Roodwater, Watchet, Somerset, England, TA23 0QL, UK

Reader Service No. 377

BIANORM® EQUILIBRIUM DIALYSIS SYSTEM

The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

BIANORM
Geräte
P.O. Box 650126
D-81215 Munich

Reader Service No. 12

Scientists moving into sales

Ann Pinn

Many scientists find bench work unrewarding after some years spent in the laboratory and if there are no prospects of promotion — especially in the research environment — they may consider moving to a technical sales career.

FAR from being "just an option", a career step for a scientist into sales should be considered very carefully indeed. It is true to say that sales is a vocation and more of a way of life than a nine-to-five job. Salespeople often work very long hours and travel extensively; they need to be lively, enthusiastic and even-tempered at all times — not always an easy task. But most important of all is an empathy with the customers they visit, and if the salesperson's background is in the laboratory a rapport and understanding can often be established quickly.

This is why scientists have the favoured background for sales of consumables and equipment into this industry; their knowledge and interest can capture the customer's attention.

The rewards in this type of career can be great and are not just monetary. It can be very satisfying to solve someone's problem, to recognize and implement a new technique or improve an old one, to persuade someone to try a new kit or purchase an instrument.

Should the scientist decide in favour of a move into sales, how to go about it? Primarily the following questions should be asked:

- Why do I want to go into sales?
- What qualities do I have to enable me to be successful?
- What preparation/homework do I need to do?

If the answer to the first question is "because I am bored in the laboratory and do not know what else to do", perhaps this

career is not suitable.

If the answer to the second is "I like meeting people", this could be a tentative start but certainly not the whole picture.

Companies are looking for individuals who relish a challenge, are self-motivated and are looking for success and recognition. For people who fit these criteria, there are various options, well known but worth reiterating:

- Contacting companies direct to see if they have any vacancies.
- Looking through the scientific journals for job opportunities.
- Contacting a recruitment consultant.

While all of these methods should be used for thoroughness, the first one can be tedious, yet a recruitment agency can make it less so by reaching a great number of companies at any one time. Also, many jobs are not advertised but registered with agencies.

An agency should be chosen that specializes in scientific and technical recruitment; a fast-moving, high-street agency may not be sympathetic to a scientist's needs, nor understand the technology involved; after all, double glazing and insurance sales undoubtedly will not appeal.

A professional agency will want to see a CV and to meet the person face-to-face, to place them in an interview situation similar to the one they will face with a future employer. The prospective salesperson can expect tough questioning and would be wise to analyse their own strengths, weaknesses and ambitions before inter-

What about money?

Remuneration and benefits packages that may be expected by a trainee sales person in the United Kingdom in their mid-20's and with a first degree (BSc/HNC) plus 2 years laboratory experience:

- £12,000—£16,000 basic guaranteed salary
- company car
- bonuses/commission (can be as little as £500 to well over £5,000 per year)
- pension/health insurance
- daily subsistence allowance
- home telephone bills/line rental
- car 'phone/fax/computer terminal

All trainees can expect full product training and a range of internal/external sales training courses.

view. A good agency does this as the first screening process of a more lengthy employment project, and if the candidate makes a strong impression, their qualities can be sold on to the company concerned. This is to be expected, as trained recruitment consultants usually have a sales and technical background, as well as an in-depth knowledge of their clients' personalities and products.

Other preparation for the scientist can include reading relevant books on the subject of sales. An agency will often provide a list of suitable publications, but a visit to the business section of a local library will unearth a wealth of information. One should bear in mind, however, that selling is all about personality, and repeating the literature's message parrot fashion will not impress.

The scientist should also arrange to spend a day with a technical sales representative; this will certainly open the eyes of the prospective sales employee and will, so to speak, kill or cure.

The satisfaction to be gained for scientists by using their technical background for a sales career can be enormous. There is no need to lose touch with new developments within science; indeed commercial people, at the sharp end, have more access to the latest technology, and, as their career grows, can often influence developments for the future.

Ann Pinn is with Delta Consultants, based at The Old Courthouse, Priory Road, St Ives, Huntingdon, Cambridgeshire PE17 4BB, UK.

